



(12)发明专利

(10)授权公告号 CN 103999126 B

(45)授权公告日 2017.02.15

(21)申请号 201180075507.8

F·蒂姆 E·巴斯 S·克莱蒙特

(22)申请日 2011.12.16

(74)专利代理机构 上海专利商标事务所有限公司 31100

(65)同一申请的已公布的文献号

申请公布号 CN 103999126 A

代理人 陈小刚

(43)申请公布日 2014.08.20

(51)Int.Cl.

G06T 7/20(2006.01)

(85)PCT国际申请进入国家阶段日

2014.06.13

(56)对比文件

(86)PCT国际申请的申请数据

PCT/EP2011/006388 2011.12.16

CN 101246602 A, 2008.08.20,

CN 101789125 A, 2010.07.28,

CN 102136139 A, 2011.07.27,

(87)PCT国际申请的公布数据

W02013/087084 DE 2013.06.20

CN 102236794 A, 2011.11.09,

W0 2010/130245 A1, 2010.11.18,

(73)专利权人 吉士特贡有限公司

审查员 杨霜雪

地址 德国吕贝克

(72)发明人 T·马丁内茨 K·埃勒斯

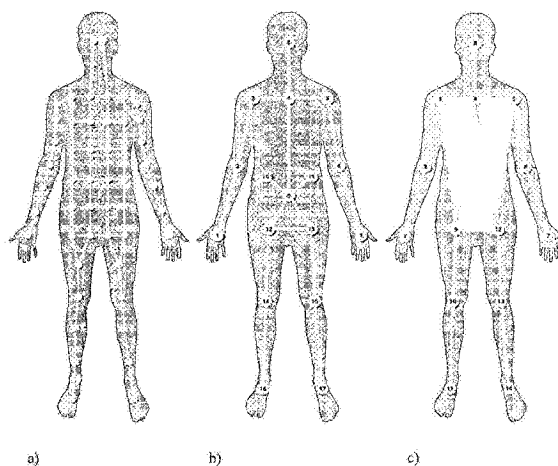
权利要求书2页 说明书12页 附图4页

(54)发明名称

用于估计姿态的方法和装置

(57)摘要

本发明涉及能够实时分析电子图像序列以供估计通过这些图像捕捉到的活动对象的姿态。本发明从骨架模型着手,该骨架模型通过3D空间中的少量节点来描述并在节点的坐标在任何时候都描述活动对象的预定部位的位置时准许对图像信息的良好数据压缩。通过在骨架模型中定义例如描述结合的对象部位或可任选的对象表面的节点对和可任选的节点三元组(它们被包含在测量出的2 1/2-D图像信息中,即对相机可见),该骨架模型同时表示该对象的先前知识。该骨架模型将被快速且准确地拟合到图像信息中。通过节点的连续位移和骨架模型的连续更新,该拟合在图像序列的两个图像之间实现。



1. 一种用于通过计算机计算骨架模型的节点的3D位置坐标的位移来对活动对象进行姿态估计的方法, 所述骨架模型被连续拟合到3D点云的序列中, 节点的3D位置坐标以表格的形式存在于电子存储器中并且根据表示所述活动对象的深度传感器相机的电子化记录的图像来确定所述3D点云,

所述骨架模型是展现出拓扑元素 N_1 个节点、 N_2 条边以及 N_3 个三角形的拓扑, 其中 $N_1, N_2 > 0$ 且 $N_3 \geq 0$, 且每一拓扑元素由节点、节点对或节点三元组来描述并且被稳定地分配给所述活动对象的一部分, 并且算术单元执行以下步骤:

- a. 随机选择所述3D点云的数据点 X ;
- b. 关于每一拓扑元素计算相对于 X 的交叉点 P 并在每一情况下标识 P 是否位于所述拓扑元素内部;
- c. 计算从 X 到每一拓扑元素的距离作为差向量 $X-P$ 的范数;
- d. 确定所有拓扑元素中展现出距 X 最小距离的拓扑元素, 且其交叉点 P 位于所述拓扑元素内部;
- e. 通过在向量 $X-P$ 的方向上将构成在步骤d确定的拓扑元素的所有节点位移, 来将所述拓扑元素位移, 节点的位移向量被乘以学习率以及权重, 所述权重从关于在步骤d确定的拓扑元素的相对于 X 的交叉点 P 得出, 以及
- f. 将步骤a到e重复预定数量的学习步骤, 同时逐渐降低所述学习率;
- g. 在 K 遍预定数量的学习步骤之后, 更新所述电子存储器的表格中的节点坐标, 其中 $K \geq 1$;
- h. 至少提供所述表格中的经更新的节点坐标以供进一步处理。

2. 如权利要求1所述的方法, 其特征在于, 在步骤b中, 关于一拓扑元素的交叉点 P 被表示为构成所述拓扑元素的各节点坐标向量的线性组合, 并且它是根据 P 是否位于所述拓扑元素的内部的表示系数来确定的。

3. 如权利要求1或2所述的方法, 其特征在于, 在步骤e中, 根据 P 的表示系数来计算所述权重。

4. 如权利要求1所述的方法, 其特征在于, 步骤a到e的重复次数在1000和5000之间。

5. 如权利要求4所述的方法, 其特征在于, 步骤a到e的重复次数在2000和3000之间。

6. 如权利要求1所述的方法, 其特征在于, 所述学习率在起始值0.5和结束值0.01之间。

7. 如权利要求1所述的方法, 其特征在于, 还包括以下步骤: 基于来自经更新的节点坐标的表格信息生成用于要被控制的装置的控制命令, 以及通过所述控制命令控制所述要被控制的装置。

8. 一种用于活动对象的姿态估计的装置, 包括:

用于检测所述活动对象的电子图像的深度传感器相机,

用于存储所述深度传感器相机的电子图像的电子存储器, 以及

被设计成在时间上与所述相机对图像进行记录步调一致地根据所述电子图像确定表示所述对象的3D点云的算术单元, 以及

其中所述存储器还存储骨架模型的节点的3D坐标的列表, 并且所述算术单元能够读出并改变各单独节点、预定节点对以及预定节点三元组的坐标作为所述骨架模型的拓扑元素的表示,

其中所述算术单元还被设计成在确定表示所述对象的3D点云之后,执行以下步骤:

- a. 随机选择所述3D点云的数据点X;
- b. 关于每一拓扑元素计算相对于X的交叉点P并在每一情况下标识P是否位于所述拓扑元素内部;
- c. 计算从X到每一拓扑元素的距离作为差向量X-P的范数;
- d. 确定所有拓扑元素中展现出距X最小距离的拓扑元素,其中所述交叉点P位于所述拓扑元素内;
- e. 通过在向量X-P的方向上将构成在步骤d确定的拓扑元素的所有节点位移,来将所述拓扑元素位移,节点的位移向量被乘以学习率以及权重,所述权重从关于在步骤d确定的拓扑元素的相对于X的交叉点P得出,以及
- f. 将步骤a到e重复预定数量的学习步骤,同时逐渐降低所述学习率;
- g. 在K遍所述预定数量的学习步骤之后,更新所述电子存储器的表格中的节点坐标,其中 $K \geq 1$;
- h. 至少提供所述表格中的经更新的节点坐标以供进一步处理。

9. 一种用于通过计算机计算骨架模型的节点的3D位置坐标的位移来对活动对象进行姿态估计的系统,所述骨架模型被连续拟合到3D点云的序列中,节点的3D位置坐标以表格的形式存在于电子存储器中并且根据表示所述活动对象的深度传感器相机的电子化记录的图像来确定所述3D点云,

所述骨架模型是展现出拓扑元素 N_1 个节点、 N_2 条边以及 N_3 个三角形的拓扑,其中 $N_1, N_2 > 0$ 且 $N_3 \geq 0$,且每一拓扑元素由节点、节点对或节点三元组来描述并且被稳定地分配给所述活动对象的一部分,并且所述系统还包括:

- a. 用于随机选择所述3D点云的数据点X的装置;
- b. 用于关于每一拓扑元素计算相对于X的交叉点P并在每一情况下标识P是否位于所述拓扑元素内部的装置;
- c. 用于计算从X到每一拓扑元素的距离作为差向量X-P的范数的装置;
- d. 用于确定所有拓扑元素中展现出距X最小距离的拓扑元素,且其交叉点P位于所述拓扑元素内部的装置;
- e. 用于通过在向量X-P的方向上将构成在步骤d确定的拓扑元素的所有节点位移,来将所述拓扑元素位移的装置,节点的位移向量被乘以学习率以及权重,所述权重从关于在步骤d确定的拓扑元素的相对于X的交叉点P得出,以及
- f. 用于将步骤a到e重复预定数量的学习步骤,同时逐渐降低所述学习率的装置;
- g. 用于在K遍预定数量的学习步骤之后,更新所述电子存储器的表格中的节点坐标的装置,其中 $K \geq 1$;
- h. 用于至少提供所述表格中的经更新的节点坐标以供进一步处理的装置。

用于估计姿态的方法和装置

发明领域

[0001] 本发明涉及用于能实时分析电子图像序列来估计在该图像序列中捕捉到的活动对象的位置和定向(即估计该活动对象的姿态)的方法和装置。本发明还涉及用软件实现本发明,并且在这一情况下涉及存储命令的计算机可读介质,该命令的执行使得根据本发明的方法被执行。

技术背景

[0002] 由计算机估计人类姿态是姿势受控的人机交互的基础。身体或手部姿势被相机捕捉,所捕捉的数字图像在计算机中被处理并解释成最终由该计算机或由它控制的装备执行的命令。在人类用户掌握了命令姿势的情况下,他不再需要单独的输入装备。

[0003] 在姿势控制的应用的特别感兴趣的领域中,在一方面在于医疗外科领域,其中外科医生想要具有对辅助设备(例如,成像设备,如超声或MRT)的直接控制,但不能用他的手触摸任何控制设备以维护无菌状态,并且在另一方面在于公共信息终端或票务机领域,当前它们仍然配备相当不卫生的触摸垫。已在商业上打开的又一应用领域是计算机游戏界。

[0004] 姿势控制方法的目的是对人的光学图像给出机器可解释的意义。这需要一种对人进行成像的装置,使得它可被电子化地评估、按照其信息内容来压缩这一图像并最终将该人的经压缩的图像转换成机器可解释的输出。该装置的输出可包括用于待控制的下游装置的控制命令。然而,它只包括馈送到下游单元以供解释这一信息的经压缩图像信息也是可能的。

[0005] 经压缩图像信息的示例是例如3D坐标系中人的右手的位置坐标的连续输出。在该过程中,只输出手部位置的单个点的坐标通常是足够的,例如在人的整个身体被成像的情况下。如果人的运动被通过图像序列进行了成像,则上述装置例如提供预定身体部位的随时间(在运动期间)变化的3D坐标。这些坐标可以用作对例如相应地控制屏幕上的光标位置的程序的可变输入。

[0006] 在图像分割期间,不能被分配给被成像的人的所有记录的图像数据(测量值)被移除,这尤其是有关背景的图像元素。这些图像元素必须从进一步评估中排除。

[0007] 使用二维数据的图像分割是困难的,特别是在用户在复杂背景前方被成像的情况下一例如,其他人在背景中运动—或者在他做出姿势(其中他的手脚朝相机运动,使得它们遮挡了他的部分躯干)的情况下。因为姿势控制要实时发生并且姿态估计通常要以25Hz或更高的视频帧率才是可能的,所以必须要在几毫秒内发生图像分割。出于这一目的,深度传感器相机可被使用,它不仅能像常规相机一样测量亮度图像,还能测量相机距对象的距离。

[0008] 一种已知的深度传感器相机被称为飞行时间相机(TOF)。它发出强度按正弦调制的红外光。按每一像素测量发出的光与该对象反射的光之间的相移。根据这一相移,可计算出光的传播时间(“飞行时间”)以及由此计算出相机距该对象点的距离。TOF相机提供与亮度图像(在TOF命名法中,通常称为振幅图像)配准的深度图。

[0009] 用于同时获得图像和距离测量值的另一种方法基于照射到要被测量的对象上并

被它反射的结构化光。相机检测到反射光一通常在与到达角度所不同的角度上,并且对由反射对象表面的位置或区域引起的投影图案的结构变化进行配准。例如,根据最初作为直线投影到对象上的、由相机捕捉到的反射线的曲率计算反射表面的凸起(这是相对于投影仪和/或相机可变的距离)是可能的。以类似的方式,通过检测点反射并确定这些点之间的距离,将各点投影在三维场景中的空间发散光束是合适的。在距投影仪更近的脸上,点距离小于图像背景中的脸上的点距离。这被用于测量脸或脸部区域距投影仪的距离。

[0010] 据此,深度传感器相机是一种除二维亮度图像之外还为每一被成像的对象点提供距离信息的装置,使得另外所有被成像的对象点沿深度轴一通常与相机的光轴相一致一的位置被测量。具有使用深度传感器相机记录的距离信息的电子图像也被称为该场景的二点五维($2^{1/2}$ D)图像。上述装置只是 $2^{1/2}$ D图像可如何产生的示例并且不一定表示最终名单。

[0011] 除了别的以外,可以从印刷出版物WO 2010/130245 A1了解 $2^{1/2}$ D图像的图像分割可如何正确地发生。图像分割根据同时测量的且按像素配准的距离值来对各相机像素检测到的亮度值进行排序。只有前景的亮度值保留在进一步评估中,假定出于改进可见性的目的,要观察的人距相机最近。前景的亮度值因而根据对该人的身体表面的成像来得到。通过本身已知的相机投影参数,被成像的对象点随后可各自被分配一组3D坐标。随后获得3D坐标的列表,其包括该人对相机而言直接可见的所有点。在3D空间中这一点“云”的内部,存在该实际的人,并且在该3D点云内部,还存在出于姿势控制的目的而需要被确定的预定人体部位的相关坐标。

[0012] 信息压缩的第二部分步骤因而可在从3D点云中确定(由图像分割来确定并表示该人)尽可能最佳地描述该人的整个姿态并适于机器解释的经缩减的一组点坐标时看到。这一步骤也被称为姿态估计。在此,姿态估计的一个目标是经缩减的数据集合的稳健性,即人类姿态的微小变化也应只造成描述该姿态的数据集合的微小变化。具体而言,描述人体部位的坐标应当尽可能远地在时间上连续的轨迹上运动,使得在任何时间都给出各坐标与这些身体部位之间的清楚相关性。

[0013] 一种已知且普遍接受的方法是定义要被尽快拟合到3D点云的人的骨架模型。

[0014] WO 2010/130245A1公开了一种能够根据 $2^{1/2}$ D图像序列进行实时姿态估计的方法,其中提出了作为节点和边的拓扑来解释的骨架模型。可被描述为节点对的边将节点之间的邻居关系结构进行编码。通过应用用于训练自组织图(“SOM”)的学习规则,节点被拟合到先前确定的点云中。

[0015] 在WO 2010/130245A1的示例性实施例中,使用44个节点和61条边的拓扑对人体的上半部建模。表示该人的3D点云包括大约6500个数据点(在真实3D空间中描绘,其中所观察的人展现了与他距相机的距离无关的所定义的大小),这些数据点中的大约10%被用于训练SOM。该拓扑的所有节点可直接被当作SOM,而指定各边可当作学习规则的特殊要求或限制。

[0016] 该拓扑是针对视频序列的每一帧分开训练的,同时一个帧的训练结果用来初始化该序列的下一帧的训练。在序列的第一帧的初始化期间,拓扑的大小通过一次性缩放优选地与在相机前方的人的大小相匹配,并且其重心位移到3D点云的重心。如果拓扑的大小曾经被正确地选择,则在该正在进行的方法期间不需要进一步适应,因为该方法是比例不变地运作的。通过应用具有以下步骤的逐模式学习规则发生对帧的训练:

[0017] a.随机选择3D点云的数据点X;

[0018] b.确定展现出距X最小距离的拓扑的节点;

[0019] c.根据拓扑的边缘规范,确定按照b确定的节点的所有邻居节点;

[0020] d.在X的方向上将按照b.和c.确定的节点位移(就此,参见WO 2010/130245A1中的式(2)和(3)),

[0021] e.将位移向量乘以学习率,该学习率与对于按照b.确定的节点相比,对于按照c.确定的节点恰好展现出一半大小(就此,请参见WO 2010/130245A1第13页第4段);

[0022] f.将步骤a到e重复预定数量的学习步骤,同时逐渐降低学习率。

[0023] 对每一帧指定最大数量的学习步骤以在预定时间间隔期间执行姿态估计一即,在这种情况下,将骨架模型拟合到3D点云并读出所有相关节点位置一是方便的。以此方式,图像序列也可以按视频帧率被分析或被更快地分析。

[0024] 虽然WO 2010/130245A1的算法很好地满足了实时姿态估计的目标,但它仍然展现出在它自己的公开部分中提到的一些弱点。具体而言,在分析其中一个人抱臂或交叉抱臂在身体前的场景时,如果各单独的节点在拓扑中被拉离它们实际的邻居很远,则学习规则可导致误解释—这可在进一步迭代的过程中被纠正。这一影响在WO 2010/130245A1中被克服,其中使用模型躯干中的锚点和学习规则的抑制节点位移离开锚点超出预定阈值的辅助条件。

[0025] WO 2010/130245 A1的教导还示出了人类关节、肩部、以及臀部的精确位置确定的困难,这在每一情况下可由若干不同的节点来表示。WO 2010/130245A1中略述的骨架模型展现出相对多的节点,在不接受相当大的姿态估计误差的情况下,其数量不能容易地降到20或更少。市场上可用于通过深度传感器相机进行姿势控制的系统已经使用根据人体解剖学设计的具有15-20个节点的骨架模型来操作。通过减少节点数,还可以获得更高的相机图像处理速度。

[0026] 解剖学推动的骨架模型另外适于借助所存储的运动模式(模板)来检测快速且复杂的运动(例如,挥动高尔夫球杆)。在这些情况下,姿势控制软件寻找检测到的姿态变化与先前存储的运动序列的最可能的匹配并使用这一已知模板来用于实际控制。这一技术已用于计算机游戏中,但它是资源密集的。最后但并非不重要是产生所存储的运动数据已经造成了相当大的成本。

[0027] 另一方面,通过SOM训练进行姿势控制完全无需模板而是仅基于运动连续性的实时检测能力。由于可被高效地实现的学习规则,它具有可靠地检测更快的人类运动并同时维持一般应用性的潜力,使得省略了软件与测量任务的可能复杂的匹配。

[0028] 技术概要

[0029] 因此,本发明的目标是压缩对象的相机图像的数字图像信息,尤其是用于姿态估计,使得可使用较少计算努力并且因而更快和/或更精确的实时执行对于对象的姿态(尤其是人类姿态)的确定。

[0030] 该目标通过独立专利权利要求的主题来实现。本发明的各有利实施例是从属专利权利要求的主题。

[0031] 本发明的一方面因而是通常按两个部分步骤来发生的信息压缩:图像分割以及姿态估计。在这一情况中,本发明尤其涉及改进姿态估计。根据本发明,提出了通过SOM训练进

行姿态估计,它能够与根据所观察的对象的解剖学来建模并展现出减少的节点数的骨架模型一起工作,它可能向预定身体部位可靠地且一致地分配每一模型节点。在此,要检测其姿态的对象的解剖学被建模成骨架模型。

[0032] 只由3D空间中少量的点(下文称为“节点”)来描述的骨架模型表示图像信息的良好信息压缩,如果节点坐标总是描述活动对象的预定部位的位置的话。同时,骨架模型表示与该对象有关的现有知识,因为例如描述包含在所测量的 $2^1/2$ D图像信息中的毗连对象部位或可任选的对象脸部的(即,相机可见的)节点对和可任选的节点三元组被定义在骨架模型中。骨架模型要被快速并精确地拟合到与该对象相对应的图像信息中。该拟合通过将节点连续地位移并与图像序列步调一致地更新骨架模型来在图像序列的两个图像之间实现。在毗连的对象部位或可能的对象脸部已经提到的情况下,假定它们作为整体来运动。根据本发明,在特定前提下,节点对或可任选的节点三元组因而同时位移。在此必须强调,下文进一步描述的位移规则不一定保留节点对各节点或可任选的节点三元组的各节点的距离,而是该位移也可造成节点对各节点或可任选的节点三元组的各节点的距离的增加。

[0033] 本发明的一实施例涉及一种用于通过计算机计算骨架模型的各节点的3D位置坐标的位移来进行活动对象(例如,人或机器人)的姿态估计的方法,其中该骨架模型被连续地拟合到3D点云序列中。节点坐标以表格的形式存在于电子存储器中并且根据来自表示运动的人的深度传感器相机的电子化记录的图像来确定3D点云。骨架模型是展现出拓朴元素 N_1 个节点、 N_2 条边以及 N_3 个三角形(其中 $N_1, N_2 > 0$ 且 $N_3 \geq 0$)的拓朴,且每一拓朴元素由节点、节点对、或节点三元组来描述并且被稳定地分配给对象的一部分(例如,人体部位或机器人的部位)。该方法由算术单元来执行并且包括以下步骤:

[0034] a. 随机选择3D点云的数据点X;

[0035] b. 关于每一拓朴元素计算相对于X的交叉点P并在每一情况下标识P是否位于该拓朴元素内部;

[0036] c. 计算从X到每一拓朴元素的距离作为差向量X-P的范数;

[0037] d. 确定所有拓朴元素中展现出距X最小距离的拓朴元素,且其交叉点P位于该拓朴元素内部;

[0038] e. 通过在向量X-P的方向上将构成在步骤d确定的拓朴元素的所有节点位移,来将该拓朴元素位移,节点的位移向量被乘以学习率以及权重,该权重从关于在步骤d确定的拓朴元素的相对于X的交叉点P得出,以及

[0039] f. 将步骤a到e重复预定数量的学习步骤,同时逐渐降低学习率;

[0040] g. 在K遍预定数量的学习步骤之后($K \geq 1$),更新电子存储器的表格中的节点坐标;

[0041] h. 至少提供该表格中的经更新的节点坐标以供进一步处理。

[0042] 在本发明的又一实施例中,在步骤b中,关于一拓朴元素的交叉点P被表示为构成该拓朴元素的各节点坐标向量的线性组合,并且它是根据P是否位于该拓朴元素的内部的表示系数来确定的。

[0043] 在本发明的又一实施例中,在步骤e中,根据P的表示系数来计算权重。

[0044] 在本发明的又一实施例中,步骤a到e的重复次数在1000和5000之间,尤其在2000和3000之间。

[0045] 有利的是,学习率可位于起始值0.5和结束值0.01之间。

[0046] 本发明的又一实施例提出了一种用于活动对象的姿态估计的装置。这一装置包括深度传感器相机、电子存储器以及可编程算术单元,该存储器存储深度传感器相机的电子图像并且算术单元被设计成在时间上与相机对图像进行记录步调一致地根据电子图像确定表示该对象的3D点云。该存储器还存储骨架模型的各节点的3D坐标的列表。该算术单元能够读出并改变各单独节点、预定节点对以及预定节点三元组的坐标来作为骨架模型的拓扑元素的表示。算术单元还被设计成在确定表示对象的3D点云之后,执行以下步骤:

[0047] a. 随机选择3D点云的数据点X;

[0048] b. 关于每一拓扑元素计算相对于X的交叉点P并在每一情况下确定P是否位于该拓扑元素内;

[0049] c. 计算从X到每一拓扑元素的距离作为差向量X-P的范数;

[0050] d. 确定所有拓扑元素中展现出距X最小距离的拓扑元素,其中交叉点P位于该拓扑元素内;

[0051] e. 通过在向量X-P的方向上将构成在步骤d确定的拓扑元素的所有节点位移,来将该拓扑元素位移,节点的位移向量被乘以学习率以及权重,该权重从关于在步骤d确定的拓扑元素的相对于X的交叉点P得出,以及

[0052] f. 将步骤a到e重复预定数量的学习步骤,同时逐渐降低学习率;

[0053] g. 在K遍预定数量的学习步骤之后($K \geq 1$),更新电子存储器的表格中的节点坐标;

[0054] h. 至少提供该表格中的经更新的节点坐标以供进一步处理。

[0055] 本发明的又一实施例涉及计算机可读存储介质,其特征在于它存储可由微处理器执行的命令,这使得微处理器执行根据本发明的所描述的实施例之一的姿态估计方法。

[0056] 附图描述

[0057] 以下参考附图使用示例性实施例更详细地描述本发明。附图中彼此对应的元素和细节已经提供了相同的参考符号。在附图中:

[0058] 图1:示出了可使用的骨架模型的草图:a)根据WO 2010/130245 1,b)来自节点和边的模型,c)来自节点、边以及三角形的模型;

[0059] 图2:示出了a)来自WO 2010/130245 A1的学习规则的图示,b)学习规则的权重的几何解释,以及c)用于边位移的情况的学习规则的效果的图示;

[0060] 图3:示出了a)学习规则的权重的几何解释以及b)三角形位移的情况下的学习规则的效果的表示;

[0061] 图4:示出了根据本发明的具有3D点云和拟合到这些点云的骨架模型的示例性图像。

[0062] 发明的详细描述

[0063] 在节点的坐标在任何时刻都描述活动对象的预定部位的位置时,只具有低节点数量的骨架模型用来高效地压缩图像信息。骨架模型是使用与该对象有关的现有知识来定义的。例如,存在与相机可见的毗连对象部位以及可任选的对象脸部有关的现有知识。骨架模型的预定元素(尤其是节点对或节点三元组)可以表示已经提及的这些对象部位或对象脸部并被作为整体拟合到对象图像中。该拟合通过将节点不断地位移并与图像序列步调一致地更新骨架模型来在图像序列的两个图像之间发生。毗连对象部位或对象脸部基本上作为整体运动,并且根据本发明,在某些前提下,节点对或可任选的节点三元组因此同时位移。

在该过程中,位移规则不一定保留节点对各节点或可任选的节点三元组的各节点相对于彼此的距离。该位移还可造成节点对各节点或可任选的节点三元组的各节点的距离的增加。对遵循保留距离的创造性的放弃简化并加速了对位移向量的必需计算并且仍然造成良好的姿态估计。将骨架模型拟合到对象图像的准确度随迭代次数(学习步骤)增加。

[0064] 以下基本上以W0 2010/130245 A1的公开内容的风格来描述本发明。要作为原则注意的是,在W0 2010/130245 A1中描述的姿态估计可被认为是起始点并且本发明的至少一个实施例可被认为是对从这一公开中已知的姿态估计方法的进一步开发。

[0065] 在这一情况中,在本发明的各实施例的以下描述中假定记录 $2^{1/2}$ D图像(单独地或作为序列)并提取表示该人的3D点云可如在W0 2010/130245 A1中描述的那样来执行。本发明尤其假定用于估计运动姿态的图像和点云可以按高于25Hz的频率来提供。

[0066] 根据本发明,现在使用受解剖学启示的骨架模型代替W0 2010/130245 A1的拓扑(在图1中被视作比较a))。图1b)中的模型是特别适合的,它将每一节点与人类在解剖学上的区别点(例如,头、肩、肘、手、臀、骨盆、膝、脚)明确地相关联。图1c)表示图1b)的模型的变型,其中躯干由三角形表示(在每一情况下由形成角的三个节点来定义)。

[0067] 通过使各节点等同于人体各部位,模型的边也被给出解剖学解释。因而,例如在图1b)中,连接节点1(右手)和2(右肘)的边一定表示右下臂。该拓扑的边因此不仅仅表示节点的邻居关系。应用来自W0 2010/130245 A1的SOM的学习规则不再能够实现该模型到在其节点数方面明显减少了的这样的骨架模型的3D点云的良好拟合,因而下文描述了适用于对受解剖学启发的骨架模型的创造性使用的姿态估计方法。

[0068] 在图2a)中,在左侧可以看到由一条边连接的两个节点 W_1 和 W_2 。点X表示从该拓扑要被拟合进的3D点云中随机选择出的点。根据W0 2010/130245A1的教导,首先,确定最接近点X的模型节点(在这种情况下是 W_1)并在朝X的方向上位移 $\varepsilon(X-W_1)$ 。在此, ε 是正实数。在本拓扑的意义上,下一邻居节点是通过边连接到 W_1 的节点 W_2 。它也在朝X的方向上位移 $\frac{\varepsilon}{2}(X-W_2)$ 。经位移的节点被标记在右侧。这一学习规则总是缩短节点之间的边。

[0069] 如果骨架模型包括如W0 2010/130245 A1中一样多的节点,则该缩短过程在作为进一步迭代和节点位移的一部分的时间过程期间再次被极大地补偿。然而,某些节点与某些身体部位的明确关联不可总是能连续的。

[0070] 因为使用图1b)或c)的解剖学拓扑,根据本发明,因而制定了总是准许拓扑的节点、边以及可任选的三角形与人的身体部位的关联的新学习规则。

[0071] 在本发明的意义上,解剖学拓扑一或即:解剖学骨架模型一至少包括:与人体各点相关联的第一数量 N_1 个节点,以及被解释为上述节点的对的选集(满足解剖学事实)的第二数量 N_2 条边。

[0072] 需要明确:在图1b)中,该模型包括节点 W_j ,其中 $j=1, \dots, 17$ 以及由节点对选集所表示的17条边。例如,对 (W_1, W_2) 、 (W_2, W_3) 或 (W_4, W_8) 是该模型的边的一部分,而例如对 (W_2, W_6) 或 (W_{10}, W_{11}) 不表示边。

[0073] 在本发明的有利设计中,第三数量 N_3 个三角形可以是解剖学模型的一部分。三角形由节点的三元组来描述,各节点指定了三角形的边。

[0074] 在图1c)中,该模型尤其包括由节点三元组 (W_3, W_4, W_9) 、 (W_9, W_4, W_{12}) 以及 (W_4, W_5, W_{12})

描述的三个三角形。

[0075] 将三角形插入解剖学模型尤其有利于对对象(例如,与身体区域相对应)进行建模,这些对象只展现出很少的内部移动性并且通常在图像中运动,使得形成三角形的边的各节点的相对位置相对于彼此只有很少变化。这例如适用于整个图像中人的躯干,但也可适用于例如在手的特写情况下的手掌或手背。姿态估计方法的用户最终总是必须自己决定要观察哪一对象或哪一拓扑看起来合适和/或特别适于他的具体目的。

[0076] 根据一实施例,本发明提供了用于来自 N_1 个节点、 N_2 条边以及 N_3 个三角形的解剖学拓扑的SOM学习规则以供拟合到3D点云,其中 N_1 、 $N_2 > 0$ 且 $N_3 \geq 0$ 。下文中,以术语拓扑元素来概括节点、边以及三角形。

[0077] SOM用拓扑的节点来标识。学习规则涉及将节点位置位移,使得拓扑拟合到3D点云中。在这之上的边和可任选的三角形的规范意味着节点的一些对和可能的一些节点三元组在每一情况下必须服从互相链接的学习规则。

[0078] 根据本发明的一实施例,学习规则是根据以下概念来形成的。从要对其执行连续姿态估计的图像序列的存在开始,作为节点位置(在3D坐标系中描述)、节点对以及可任选的节点三元组的列表的骨架模型在每一情况下在该序列的新图像存在时被更新。新图像一旦被深度传感器相机记录并且通过图像分割和投影到所观察的对象的表面的各点一例如,整个人、他的躯干、他的手,等等一的3D坐标的列表(3D点云)已被转换,SOM训练就发生,因为该点云的单独的点是随机选择的并且最接近这一点的拓扑元素在该所选点的方向上位移。位移通过3D空间中的向量加法来发生,其结果是上述节点位置的列表中各单独的节点位置被改变或更新。

[0079] 在这一位移之后,3D点云的下一点是随机选择的,并且最接近现在被选择的点的拓扑元素(非常有可能是不同的一个)朝该点位移。点选择和位移被重复预定数量的步骤,总体位移距离随每一步骤而降低。在足够高的预定数量的步骤之后,骨架模型针对新图像被最终更新。

[0080] 优选地,执行拓扑元素的至少1000和至多5000个位移来将骨架模型拟合到3D点云中。尤其优选的是,位移的数量在2000和3000之间。节点位置的若干1000个位移可使用当今的计算机在几毫秒内实现。

[0081] 根据这一点,最接近所选择的点的拓扑元素可能不是节点。通过将以互相链接的方式来确定边或三角形的所有节点位移来对边或三角形进行位移。取决于要位移的拓扑元素,位移也可同时考虑一个、两个或三个节点。

[0082] 最接近拓扑元素的确定在下文更详细地解释,并且指定了拓扑元素的特定位移规则。

[0083] 首先,为3D点云的点X首先确定X在所有拓扑元素方面的交叉点—换言之,对于3D空间中的坐标点,在使用深度传感器相机对所观察的对象成像以及后续的图像分割和投影之后,表示该人的身体表面的点并且从所有这些点中随机选择。关于一拓扑元素,X的交叉点P是由该拓扑元素形成的3D空间的子空间的最接近点X的点。在该过程中,各节点形成只包含该节点自身的零维子空间。各边形成3D空间中的贯穿限定该边的节点的直线。三角形形成3D空间中包含限定该三角形的节点的平面。

[0084] 由一拓扑元素形成的子空间的最接近点X的点是使用距离测量来计算的。3D空间

中的距离也可以使用任何范数来确定。优选地,使用欧几里得范数(也称为 L_2 范数或毕达哥拉斯距离),但也可使用其他距离测量。

[0085] 如果拓扑元素是节点,则交叉点P处于节点中,并且在此它恰好与这一节点重合。

[0086] 如果拓扑元素是边,则交叉点P位于3D空间中的直线上。

[0087] (1) $P=W+\alpha \Delta W$

[0088] 在此,W表示该边的一任何一第一节点, ΔW 表示该边的第二节点和第一节点之间的差向量,且 α 是实数。需要明确:该边由节点对(W,W+ ΔW)来描述。

[0089] 如果拓扑元素是三角形,则交叉点P位于3D空间中的平面上。点P可被表示为:

[0090] (2) $P=W+\sigma_1 \Delta W_1+\sigma_2 \Delta W_2$

[0091] 其中W是三角形的一任何一第一节点, ΔW_1 、 ΔW_2 是该三角形的第二节点和/或第三节点与第一节点之间的差向量,并且 σ_1 、 σ_2 是实数。需要明确:三角形由节点三元组(W,W+ ΔW_1 ,W+ ΔW_2)来描述。

[0092] 式(1)和(2)的系数 α 、 σ_1 、 σ_2 应当在下文指定为交叉点的“拓扑顺从表示系数”。形式上,它们是向量P相对于3D空间的子空间的非标准化且可任选地斜角底边的分量,该子空间由其确定P的拓扑元素来形成。关于拓扑元素的交叉点P由确定该拓扑元素的节点坐标向量的线性组合来表示也是可能的。

[0093] 随后检查点X关于拓扑元素的交叉点是否位于这些拓扑元素内部。

[0094] 根据定义,关于节点的所有交叉点都位于这些节点内部。同样,根据定义,关于节点的交叉点的拓扑顺从表示系数总是1。

[0095] 如果关于边的交叉点位于定义该边的节点之间,则该交叉点恰好处于该边内部。这意味着当且仅当交叉点的表示系数 α 在0和1之间时,交叉点才位于边的内部。

[0096] 如果关于三角形的交叉点在由节点限定的三角形表面的内部,则该交叉点恰好处于该三角形内部。对于关于三角形的交叉点,当且仅当它的表示系数 σ_1 、 σ_2 以及它们的和 $\sigma_1+\sigma_2$ 在0和1之间时才是这种情况。

[0097] 3D点云的点X距一拓扑元素的距离可作为X与交叉点P之间的差向量D的范数来计算,即在下式中 $D:=X-P$ 。优选地,使用欧几里得范数。以此方式,距离

[0098] (3) $d=\|D\|_2=\|X-P\|_2=\sqrt{(x_1-p_1)^2+(x_2-p_2)^2+(x_3-p_3)^2}$

[0099] 其中为拓扑元素中的每一个来计算作为向量X和P关于常规3D坐标系(其中描述了3D点云和骨架模型)的实部的 x_1 、 x_2 、 x_3 、 p_1 、 p_2 、 p_3 。

[0100] 点X距一拓扑元素的距离进一步只在关于这一拓扑元素的交叉点P处于该拓扑元素内部的情况下使用。如果情况并非如此,则该交叉点被丢弃,并且不计算距离或计算得到的距离随后被忽略。

[0101] 这一选择确保在进一步处理期间只有实际上服务于拟合骨架模型目的的那些距离才被考虑。点X可例如恰好处于3D空间中的由表示该人的躯干的三角形形成的平面中。交叉点P随后与X重合,并且距离d是零。然而,如果同时,该点位于表示右手的节点的位置的邻近区中,并且右臂远离身体伸出,则位移规则要被应用于该节点而非躯干的三角形,即使它的平面距X的数字距离小于该节点距X的距离。

[0102] 根据所有计算得到的距离—以及估计中保留的其他距离,通过寻找最小值,标识

出最接近点X(即展现出距X的最小距离)的拓扑元素。

[0103] 根据本发明,所标识的拓扑元素现在被位移,限定该拓扑元素的节点可能被联合位移。

[0104] 如果最接近点X的拓扑元素是节点,则位移根据下式发生

$$[0105] \quad (4) \quad W \xrightarrow{\text{更新}} W + \varepsilon_t D$$

[0106] 这在本领域中是已知的。在WO 2010/130245 A1中—在此的式(4),同样,学习率

$$[0107] \quad (5) \quad \varepsilon_t = \varepsilon_i \left(\frac{\varepsilon_f}{\varepsilon_i} \right)^{\frac{t}{t_{\max}}}$$

[0108] (其是迭代步骤的函数)被解释。在此, ε_i 和 ε_f 是学习率的预定起始和结束值,并且 $t_{\max}$ 是学习步骤的预定最大数量。运行指数t将学习步骤(迭代)计数上至 $t_{\max}$ 。对于3D点云的每一新选择的点X,分子t加1,直至它达到 $t_{\max}$ 。学习率的优选条件是 $\varepsilon_i=0.5$ 且 $\varepsilon_f=0.01$ 。

[0109] 如果最接近点X的拓扑元素是边,则位移根据下式发生

$$[0110] \quad (6) \quad W \xrightarrow{\text{更新}} W + \varepsilon_t (1 - \alpha) D$$

$$[0111] \quad (7) \quad W + \Delta W \xrightarrow{\text{更新}} W + \Delta W + \varepsilon_t \alpha D$$

[0112] 在图2b)中,示出了边(W, W+ΔW)、3D点云的点X,以及它关于该边的相关联的交叉点P以及差向量D=X-P。拓扑顺从表示系数α将被认作边长度 $\|\Delta W\|_2$ 的分数,据此必须从节点W开始在该方向上走动ΔW才能到达点P。很明显,在图2b)中是 $0 < \alpha < 1$ 的情况,并且交叉点因而在边的内部。

[0113] 图2c)概括示出了这两个节点的位移。两者都在向量D=X-P的方向上运动,即,-与现有技术不同-并非直接朝X运动。如果与节点W+ΔW相比,点X更接近节点W,则这也适用于交叉点P,并且因此 $\alpha < 0.5$ 。节点W随后比节点W+ΔW位移得更远。以此方式,整条边靠近根据X和/或P距节点距离来加权的点X。

[0114] 如果最接近点X的拓扑元素是三角形,则位移根据下式发生

$$[0115] \quad (8) \quad W \xrightarrow{\text{更新}} W + \varepsilon_t (1 - \delta_1 - \delta_2) D$$

$$[0116] \quad (9) \quad W + \Delta W_1 \xrightarrow{\text{更新}} W + \Delta W_1 + \varepsilon_t \delta_1 D$$

$$[0117] \quad (10) \quad W + \Delta W_2 \xrightarrow{\text{更新}} W + \Delta W_2 + \varepsilon_t \delta_2 D$$

[0118] 图3示出了三角形的位移规则。

[0119] 图3a)示出了骨架模型的起始三角形和随机选择的点X。与X相关联的交叉点P处于三角形平面中并且可根据式(2)被表示为节点坐标向量的线性组合。出于说明的目的,表示系数 σ_1 、 σ_2 被记在三角形的侧边处。它们要与图2b)中的系数α类似地解释为三角形侧边的长度的分数。交叉点处于三角形内部,所以X距该三角形的距离被确定为向量D=X-P的范数并且在进一步计算中被使用。如果这一距离结果是X距骨架模型的所有拓扑元素的最小距离,则该三角形被位移。

[0120] 经位移的三角形在图3b)中概括示出。在此,同样,所有节点—与三角形的各侧上的以及三角形的表面中的所有点一起—被位移,相应地用节点距点X和P的初始距离来加权。使用距离的这一加权对姿态估计的效率而言是重要的,因为借此避免了不必要的误差。

这一优点可容易理解,尤其在查看图3b)时:如果点P非常接近构成该三角形的三个节点之一,则该三角形被位移,使得几乎只有该最接近节点被运动而另两个节点基本上维持它们的位置。因而在此所描述的方法中,存在节点、边以及三角形位移之间的“平滑过渡”。

[0121] 从图2c)可以容易地认识到,边($W, W + \Delta W$)决不会通过应用学习规则(6)和(7)而缩短,而是非常可能被延长。在应用式(8)到(10)的学习规则时,这对三角形($W, W + \Delta W_1, W + \Delta W_2$)的侧边而言也成立。

[0122] 在本发明的又一实施例中,为了在训练过程期间不会获得失去控制的边长度和三角形侧边长度,引入了“收缩参数” δ 。它可以例如被设为 $\delta = 0.05$ 。同时,学习规则被修改,使得在位移期间,节点朝彼此稍微运动。

[0123] 代替规则(6)和(7),优选地

[0124] (11) $W \xrightarrow{\text{更新}} W + \delta \Delta W + \varepsilon_t (1 - \alpha) D$

[0125] (12) $W + \Delta W \xrightarrow{\text{更新}} W + (1 - \alpha) \Delta W + \varepsilon_t \alpha D$

[0126] 被使用,并且代替规则(8)到(10)的式,

[0127] (13) $W \xrightarrow{\text{更新}} W + \delta \Delta W_1 + \delta \Delta W_2 + \varepsilon_t (1 - \delta_1 - \delta_2) D$

[0128] (14) $W + \Delta W_1 \xrightarrow{\text{更新}} W + (1 - \delta) \Delta W_1 + \delta (\Delta W_2 - \Delta W_1) + \varepsilon_t \delta_1 D$

[0129] (15) $W + \Delta W_2 \xrightarrow{\text{更新}} W + (1 - \delta) \Delta W_2 + \delta (\Delta W_1 - \Delta W_2) + \varepsilon_t \delta_2 D$

[0130] 被使用。

[0131] 所描述的方法可被用于通过计算根据如图1中的人体解剖学建模的骨架模型的节点的位移来估计人类姿态。然而,同样明显的是,相同方法可同样应用于动物或活动机器人的运动。本发明包括由互连且可相对于彼此运动的各部分构成的所有对象的姿态估计,并且对于这些部分,来自节点、边以及可任选的三角形的骨架模型可被方便地定义。

[0132] 骨架模型是具有拓扑的节点的3D坐标的表格和具有用于构成边和三角形的预定节点对以及可任选地节点三元组的列表。在姿态估计的执行期间,节点对和节点三元组的列表不能被改变,只有节点坐标可以改变。具有节点坐标的表格可在电子存储器中获得并且可由算术单元读出。算术单元使用同样存储的来自深度传感器相机的图像和从中确定的3D点云,相对于骨架模型到点云的连续拟合来确定节点位置的位移。

[0133] 根据本发明的一实施例发生将解剖学模型拟合到点云中—因此训练SOM—,概括如下:

[0134] a. 随机选择3D点云的数据点X;

[0135] b. 关于一拓扑元素计算相对于X的交叉点P,并确定它的至少一个拓扑顺从表示系数;

[0136] c. 如果交叉点不位于该拓扑元素内部则丢弃它;

[0137] d. 计算X到该拓扑元素的距离作为差向量X-P的范数;

[0138] e. 对于解剖学模型的所有拓扑元素,重复步骤b.到d.;

[0139] f. 确定展现出距X最小距离的拓扑元素;

[0140] g. 通过将构成按照f.确定的拓扑元素的所有节点在向量X-P的方向上位移,来将该拓扑元素位移,该位移向量被乘以学习率和权重,该权重从关于按照e.确定的拓扑元素

的相对于X的交叉点P的拓扑顺从表示系数得出,以及

[0141] h.将步骤a.到g.重复预定数量的学习步骤,同时逐渐降低学习率。

[0142] 如何从拓扑顺从表示系数中获得在步骤g中提及的权重确切方式可以从式(6)到(10)了解或作为替换从式(11)到(15)了解,后一组包括附加的预定参数 δ 。根据定义,根据式(4),作为节点的拓扑元素的位移的权重是1。

[0143] 如果预定数量的学习步骤已经达到,则算术单元已经为骨架模型的所有节点计算了经位移的坐标。作为规则,经位移的坐标将偏离所有节点的初始制成表格的节点坐标,但在例外情况下,对于个别节点,经位移的坐标也可以与其初始制成表格的节点坐标相同。

[0144] 经位移的节点坐标被算术单元写入电子存储器中,初始坐标条目或者i)被重写或者ii)被指定为过期并且在进一步学习步骤中不再被使用。以此方式,节点坐标在电子存储器的表格中被更新。有利地,以下是可能的:执行在只有一遍预定数量的学习步骤之后更新电子存储器的表格中的节点坐标的步骤并随后使用来自深度传感器相机的图像序列的后续图像和从中确定的点云来开始步骤a到h的下一遍。

[0145] 同样,以下也是可能的:对于一个且相同的点云将各学习步骤从头到尾相继循环若干次来计算节点坐标的多个位移。该多个位移随后可例如在算术上取平均,并且节点坐标只与经平均的位移一起在电子存储器的表格中被更新。这一过程在计算方面更加复杂并且因而更慢,但也可有效地抑制任何偶然发生的不利的位移(例如,在点云也包括随机选择的具有外部位置的点的情况下)。

[0146] 广义而言,更新节点坐标因此在K遍预定数量的学习步骤之后发生,其中 $K \geq 1$ 。

[0147] 在图4中使用来自视频序列的示例性图像呈现了使用来自图1b)的骨架模型的创造性的姿态估计方法的一示例性实施例。深度传感器相机的视频图像使用本质上已知的图像分割来不断地提供表示运动的人的3D点云。使用上述学习规则,解剖学骨架模型被实时拟合,并且该模型节点与人的不同身体部位的关联始终保持正确,如可从图4清楚地了解的。

[0148] 现在实现的稳健的关联节点-身体部位准许通过姿势控制的显著更稳定的命令输入,例如简单地通过跟踪右手的运动。解释骨架-节点位置的机器可以忽略(在必要时)其余模型节点或还将它们分类成附加输入。因而,例如以下是可能的:通过同时并分开地跟踪和解释右手和左手的运动来执行更复杂的输入。将左手举起到头的高度可例如等效于按下键盘上的某一键(例如,换挡键),或检索用于右手的输入的另一组命令解释。

[0149] 除了通过深度传感器相机来快速检测并提供 $2^1/2$ D图像之外,所描述的方法还需要至少用于临时存储测量数据并且用于存储解剖学骨架模型的拓扑元素的位置坐标的电子存储器,以及还需要与该存储器通信的执行所描述的计算并且具体而言促使存储器中的拓扑元素的位置坐标的连续更新的电子算术单元。此外,相同算术单元可以在时间上保持步调一致地独自解释相应的当前骨架模型并且例如将它转换成用于从属装置的控制命令,或者可能只输出所选节点位置以用于进一步解释。

[0150] 如在开头描述的,所描述的用于姿态估计的方法可以由一种装置(具体而言是算术单元)来执行。算术单元可以是市场上可购得的可编程微处理器,但它也可能使用FPGA或ASIC。本发明的又一实施例涉及存储可由算术单元执行并且使得算术单元执行在此描述的用于姿态估计的目的的计算的命令的存储介质。

[0151] 包括用于确定距离坐标的算术单元的深度传感器相机可尤其也被直接设计成执行本发明方法。作为具有根据本发明设计的至少一个算术单元的构造单元的此类相机相应地适于将运动的人的图像直接转换成他的基本身体部位的3D坐标。这可媲美运动捕捉装置,然而可省却迄今为止仍然常见的在人体上的标记。

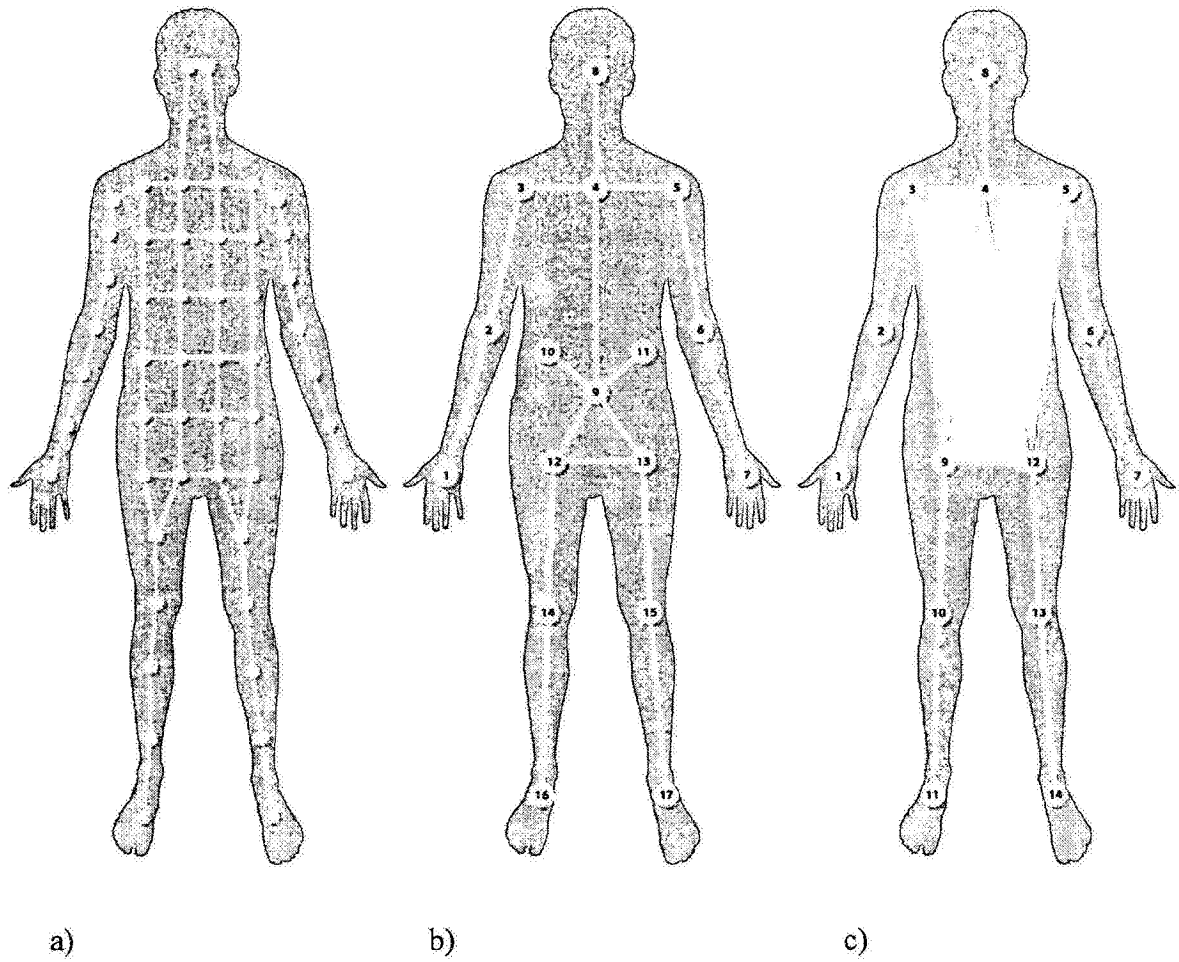


图1

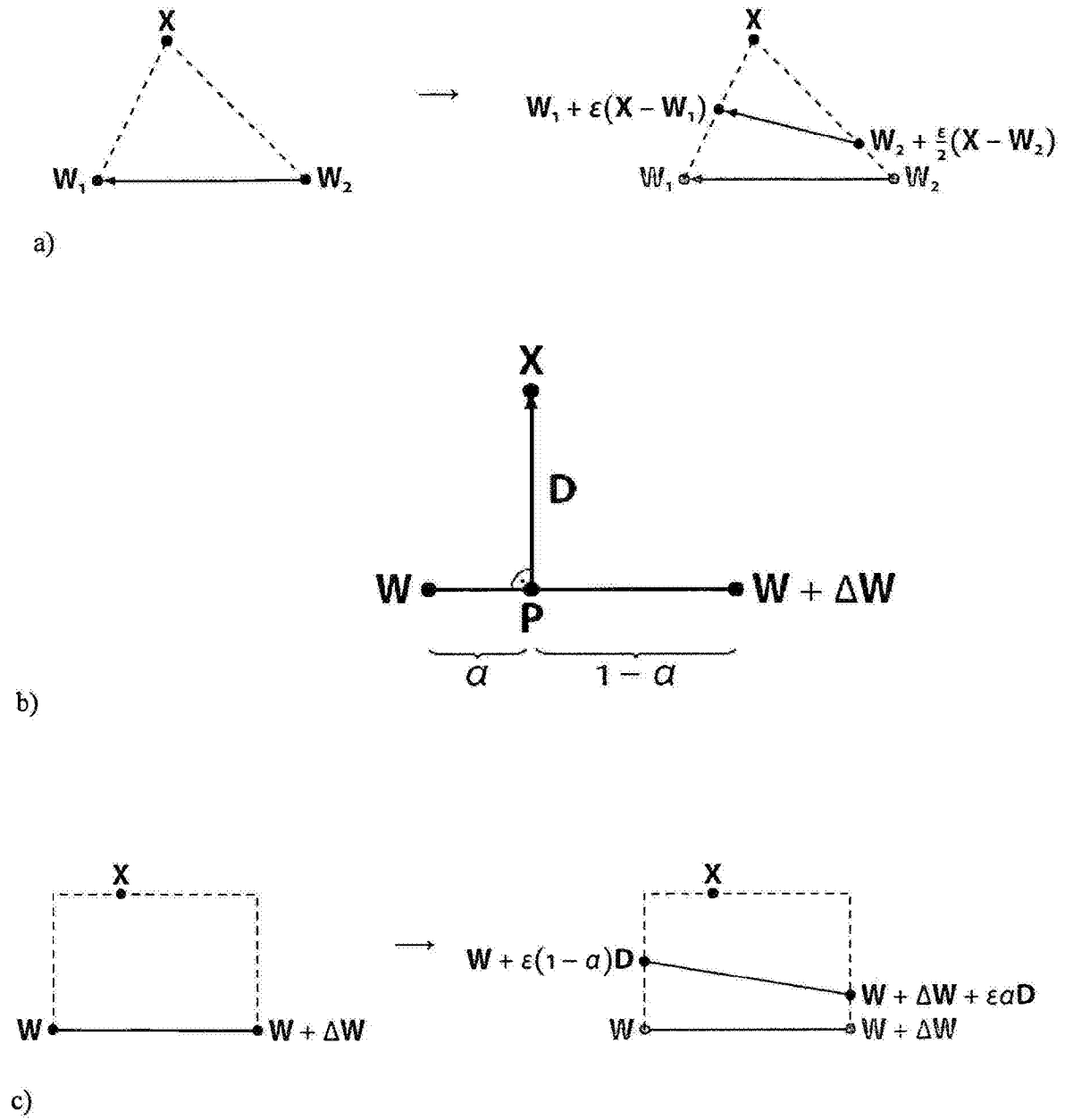
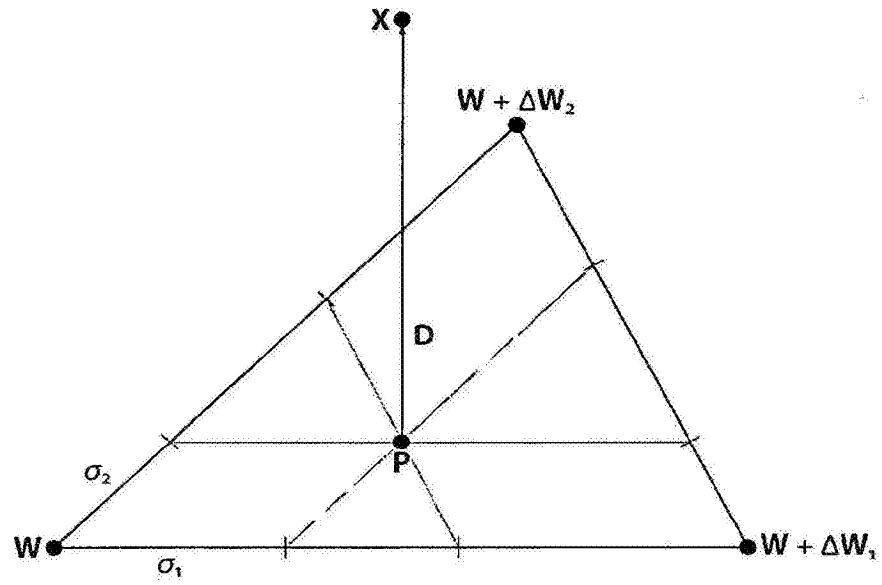
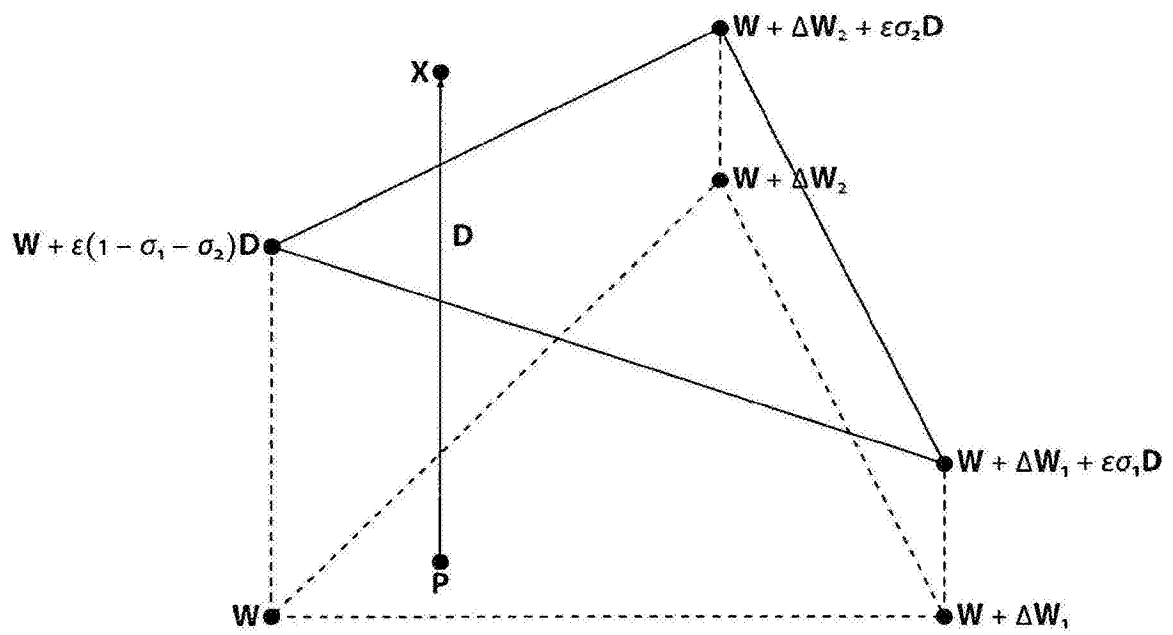


图2



a)



b)

图3

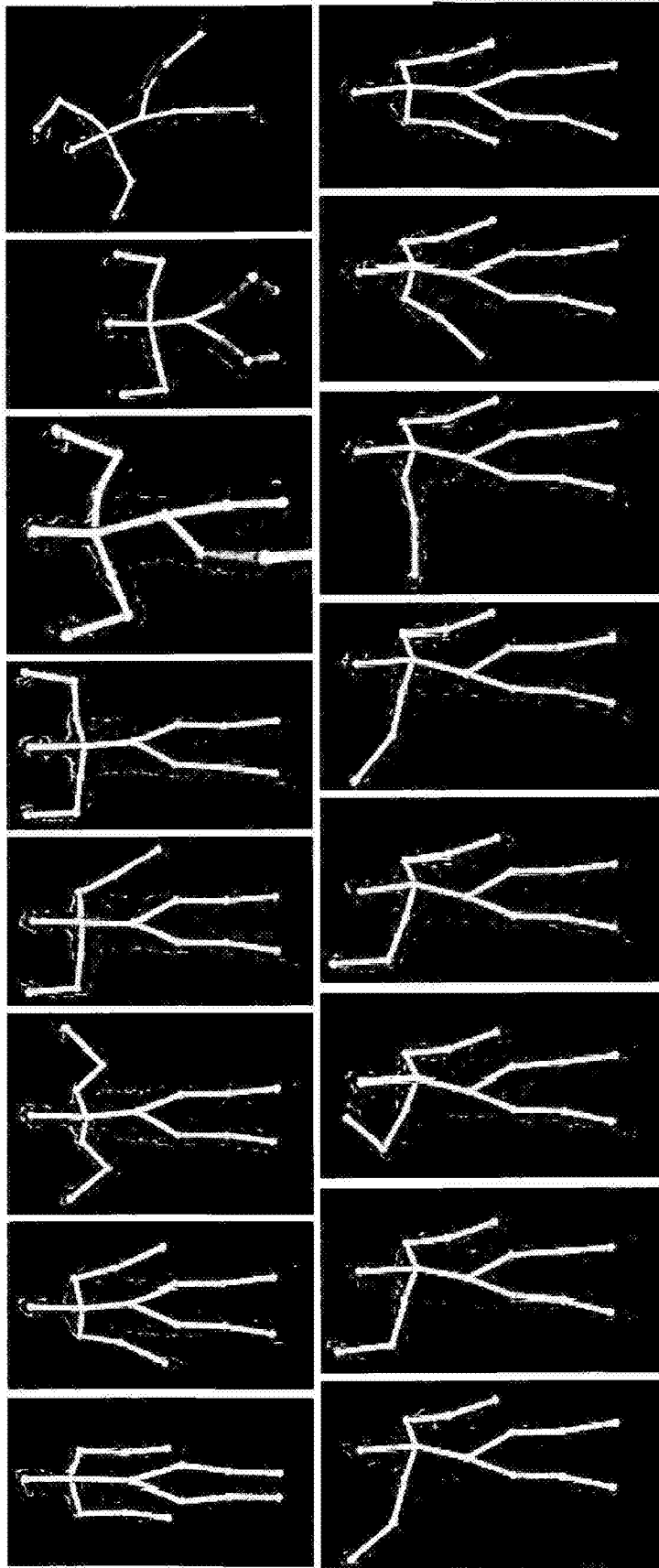


图4