



## (12)发明专利

(10)授权公告号 CN 104268138 B

(45)授权公告日 2017.08.15

(21)申请号 201410205213.8

(22)申请日 2014.05.15

(65)同一申请的已公布的文献号

申请公布号 CN 104268138 A

(43)申请公布日 2015.01.07

(73)专利权人 西安工业大学

地址 710032 陕西省西安市未央区学府中路2号

(72)发明人 肖秦琨 谢艳梅

(74)专利代理机构 西安新思维专利商标事务所  
有限公司 61114

代理人 黄秦芳

(51)Int.Cl.

G06F 17/30(2006.01)

G06K 9/00(2006.01)

(56)对比文件

CN 103294996 A, 2013.09.11,

Feng zhou等.canonical time warping for alignment of human behavior.《Carnegie mellon university》.2009,

陈炉军.动态深度数据匹配及其应用.《中国优秀硕士学位论文全文数据库 信息科技辑》.2014,

陈炉军.动态深度数据匹配及其应用.《中国优秀硕士学位论文全文数据库 信息科技辑》.2014,

罗鸣.基于Kinect传感器的骨骼定位研究.《中国优秀硕士学位论文全文数据库 信息科技辑》.2014,

wang xue等.Human-like character animation of maize driven by motion capture data.《information and computational science》.2011,

审查员 王芬

权利要求书3页 说明书7页 附图3页

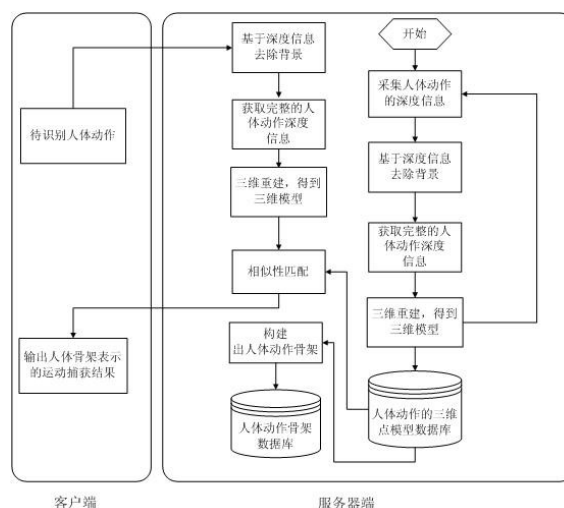
(54)发明名称

融合深度图和三维模型的人体运动捕捉方法

法

(57)摘要

本发明涉及一种融合深度图和三维模型的人体运动捕捉方法。光学式运动捕获方法需粘贴标记点,交互不便,标记点易混淆、遮挡。本发明采集人体动作的深度信息,去除运动目标背景,获取完整的人体动作深度信息,转化成人体的三维点云信息,获得人体动作三维模型,建立数据库,与建立的人体动作骨架数据库数据一一对应,提取待识别人体动作的深度信息构建三维模型,然后与三维模型数据库中的人体动作进行相似性匹配,按相似性排序输出人体动作骨架作为运动捕捉结果。本发明无需在人体上安装传感器或添加标记点,方便易实现,在运动序列匹配上使用正则时间规整的方法,提高了两个序列匹配的精度,使匹配的时间大幅降,保证了运动捕捉的速度和精度。



1.融合深度图和三维模型的人体运动捕捉方法,其特征在于:

由以下步骤实现:

步骤一:采集人体动作的深度信息,去除运动目标背景,获取完整的人体动作深度信息;

步骤二:将提取的人体动作深度信息信息转化成人体的三维点云信息,对人体进行三维重建,获得人体动作的三维模型;

步骤三:重复步骤一和步骤二,基于大量采集的人体动作深度信息,建立人体动作的三维模型数据库 $M = \{Y_1, Y_2, \dots, Y_n\}$ ,  $M$ 表示三维模型数据库,  $Y_n$ 表示第 $n$ 个三维模型;

步骤四:根据人体结构及人体动作的三维模型构建出人体动作骨架,建立人体动作的骨架数据库 $G = \{S_1, S_2, \dots, S_n\}$ ,  $G$ 表示人体骨架数据库,其中 $S_n$ 表示第 $n$ 个人体骨架数据,其中人体动作的骨架数据与三维模型数据一一对应;

步骤五:提取待识别人体动作的深度信息,基于深度信息构建待识别人体动作的三维模型,然后与三维模型数据库中的人体动作进行相似性匹配,按相似性排序输出人体动作骨架,相似性距离值最小的骨架作为最优骨架,以最优骨架作为运动捕获结果;

步骤一中,基于采集的人体动作深度信息,去除运动目标背景,获取完整的人体动作深度信息的具体步骤为:

(1) 采集的一个人体动作深度图用 $F(x, y, d_p)$ 表示,其中, $x$ 、 $y$ 分别为像素坐标系下的横坐标和纵坐标,  $d_p$ 为深度信息;假设基于深度信息分割背景区域和目标区域的阈值为:

$$T = (\maxDepthVulue + \minDepthVulue) / 2$$

其中,  $\maxDepthVulue$ 和 $\minDepthVulue$ 分别为图像深度值的最大和最小值,将 $T$ 记录在 $T_0$ 中,根据阈值 $T$ ,将 $F(x, y, d_p)$ 分为背景区域和目标区域,分别求出背景区域和目标区域的平均深度值 $du_1$ 和 $du_2$ ;

(2) 重新计算 $T = (du_1 + du_2) / 2$ ,判断 $T$ 与 $T_0$ 是否相等,如果不相等,那么将 $T$ 记录在 $T_0$ 里,重复上述步骤,直到 $T = T_0$ 成立,终止算法;把最后得到的 $T$ 作为最佳阈值对 $F(x, y, d_p)$ 进行分割,去除背景,得到完整的人体动作深度信息 $d_0$ ;

步骤二中,将提取的人体动作深度信息转化成人体的三维点云信息,对人体进行三维重建,获得人体动作的三维模型的具体步骤为:

(1) 将提取出的深度信息 $d_0$ 归一化,假设深度值的个数为 $N$ ,计算出深度值中的最大值 $\max d_0(k)$ 和最小值 $\min d_0(k)$ ,归一化后的深度值为:

$$z(n) = (d_0(n) - \min d_0(k)) / (\max d_0(k) - \min d_0(k)), n = 1, 2, \dots, N$$

(2) 令 $Z = z(n)$ ,世界坐标系是以摄像机为原点的,经过标定后摄像机可以看做是理想的成像模型,根据简单的相似三角形变换原理:

$$\frac{X}{x} = Z \times \frac{1}{f}$$

$$\frac{Y}{y} = Z \times \frac{1}{f}$$

从中可以计算得到世界坐标系的 $X$ 、 $Y$ 值,其中, $x$ 、 $y$ 分别为像素坐标系的横坐标和纵坐

标,  $f$  为摄像机的焦距; 最终可以得到人体动作的三维点云信息  $(X, Y, Z)$ ;

步骤四中, 根据人体结构及人体动作的三维模型构建出人体动作骨架的具体步骤为:

(1) 在深度图中将身体躯干近似视为四边形, 记为  $Q$ ;

$Q$  上方的两个顶点就是肩关节  $a_1$  和  $a_2$  的位置;

颈关节  $b$  为  $a_1$ 、 $a_2$  连线的中点;

沿  $b$  点向上, 最上端就是头部节点  $c$  的位置;

$Q$  下方的两个顶点为髋关节  $d_1$  和  $d_2$  的位置;

胯关节  $e$  在  $d_1$ 、 $d_2$  的中点位置;

(2) 确定手关节和肘关节位置从  $a_1$ 、 $a_2$  处开始进行搜索, 若手臂伸直, 根据上下手臂长度比例确定肘关节  $f_1$ 、 $f_2$  和手关节  $g_1$ 、 $g_2$  的位置; 若手臂弯曲, 在肘关节的位置有拐点, 拐点的位置为  $f_1$ 、 $f_2$ , 沿  $f_1$ 、 $f_2$  继续搜索至终点为手关节  $g_1$ 、 $g_2$  的位置;

同理, 根据手关节和肘关节位置的确定方法可以确定出膝关节  $h_1$ 、 $h_2$  和脚关节  $i_1$ 、 $i_2$  的位置;

(3) 将各个关节坐标按人体结构顺序用直线相连接, 就得到了人体动作骨架;

步骤五中, 提取待识别人体动作的深度信息, 基于深度信息构建待识别人体动作的三维模型, 然后与三维模型数据库中的人体动作进行相似性匹配的具体步骤为:

(1) 将三维模型表示的人体动作层次聚类算法进行聚类:

① 假设一个参数  $m$  表示最终聚类簇的数目, 将每个输入的数据作为单独的一个数据簇  $D_0$ , 求出与每个  $D$  相邻的最近的簇  $D_x$ , 数据簇间的距离可以看做是数据簇中心间的距离;

② 将距离最近的两个数据簇  $D_p$  和  $D_q$  进行合并, 产生的新的数据簇  $D_n$ , 再计算  $D_n$  与其它簇的距离, 如果当前数据簇的数目大于  $m$ , 那么就转到②继续进行数据簇的合并, 否则算法结束, 得到聚类结果  $D = \{u_1, u_2, \dots, u_m\}$ , 其中  $u$  表示每个类的值;

(2) 计算待识别人体动作序列各个动作帧与三维模型数据库中的人体动作序列各个动作帧之间的距离:

① 假设待识别人体动作和数据库中任意一个三维模型聚类结果分别表示为:

$$D_i = \{u'_1, u'_2, \dots, u'_m\}$$

$$D_j = \{v_1, v_2, \dots, v_m\}$$

从聚类树的根节点开始, 进行深度优先搜索直至叶节点;

② 在遍历搜索的过程中对  $D_i$  和  $D_j$  中各个类的对象计算距离, 将各个类的距离求和, 得到两颗聚类树的欧式距离:

$$dist = \sum_{\substack{i=1 \dots m \\ j=1 \dots m}} \sqrt{(u'_{ix} - v_{jx})^2 + (u'_{iy} - v_{jy})^2 + (u'_{iz} - v_{jz})^2}$$

(3) 用正则时间规整的方法计算出最优匹配路径, 得到最优匹配序列:

① 假设待识别的三维模型动作序列为  $\mathbf{X} = [x_1, \dots, x_m] \in R^{d_x \times m}$ , 三维模型数据库中的一组动作序列为  $\mathbf{Y} = [y_1, \dots, y_n] \in R^{d_y \times n}$ ,  $m$  和  $n$  分别表示两个动作序列的长度, 两个序列中各个动作帧之间的距离用上述方法可以计算得到;

②因为m和n可能不相等,所以其对应的匹配关系可能有很多种,利用正则时间规整的方法,计算出序列间的最小距离J,同时获得最佳匹配路径 $P_x$ 和 $P_y$ :

$$\min_{\{\mathbf{V}_x, \mathbf{V}_y\} \in \Phi, \{\mathbf{p}_x, \mathbf{p}_y\} \in \Psi} J_{ctw} = \left\| \mathbf{V}_x^T \mathbf{X} \mathbf{W}_x - \mathbf{V}_y^T \mathbf{Y} \mathbf{W}_y \right\|_F^2 + \phi(\mathbf{V}_x) + \phi(\mathbf{V}_y)$$

这里,  $\mathbf{W}_x = \mathbf{W}(p_x) \in \{0, 1\}^{m \times 1}$  和  $\mathbf{W}_y = \mathbf{W}(p_y) \in \{0, 1\}^{n \times 1}$  是两个二进制的规整矩阵, 对每一步的  $t \in \{1:1\}$  有  $w_{pt,t} = 1$ , 其他情况时  $w_{pt,t} = 0$ , 其中1是正则时间规整算法自动选择的需要匹配的动作帧的数目,  $1 \geq \max(m, n)$ ;  $\phi(\cdot)$  是正则项:

$$\phi(\mathbf{V}) = \frac{\lambda}{1 - \lambda} \|\mathbf{V}\|_F^2$$

$P_x$ 和 $P_y$ 表示所有可能的匹配路径, 需要满足的约束条件为:

$$\Psi = \{ \{P_x, P_y\} \mid P_x \in (1:m)^1, P_y \in (1:n)^1 \}$$

$$\text{边界: } [p_1^x, p_1^y] = [1, 1], [p_l^x, p_l^y] = [m, n]$$

$$\text{单调性: } t_1 \geq t_2 \Rightarrow p_{t_1}^x \geq p_{t_2}^x, p_{t_1}^y \geq p_{t_2}^y$$

$$\text{连续性: } [p_t^x, p_t^y] - [p_{t-1}^x, p_{t-1}^y] \in \{[0, 1], [1, 0], [1, 1]\}$$

$\mathbf{V}_x \in R^{d_x \times d}$  和  $\mathbf{V}_y \in R^{d_y \times d}$  是线性变化矩阵,  $d \leq \min(d_x, d_y)$ , 需满足的约束条件为:

$$\Phi = \left\{ \left\{ \mathbf{V}_x, \mathbf{V}_y \right\} \left| \begin{aligned} \mathbf{V}_x^T \left( (1 - \lambda) \mathbf{X} \mathbf{W}_x \mathbf{W}_x^T \mathbf{X}^T + \lambda \mathbf{I} \right) \mathbf{V}_x &= \mathbf{I}, \\ \mathbf{V}_y^T \left( (1 - \lambda) \mathbf{Y} \mathbf{W}_y \mathbf{W}_y^T \mathbf{Y}^T + \lambda \mathbf{I} \right) \mathbf{V}_y &= \mathbf{I} \end{aligned} \right. \right\}$$

其中,  $\lambda \in [0, 1]$ ;

③重复以上步骤, 利用正则时间规整方法, 计算三维模型数据库中每一组动作序列Y分别与X的最小距离J及最佳匹配路径 $P_x$ 和 $P_y$ ; 然后求出各个J的最小值和其对应的 $P_x$ 和 $P_y$ , 从而获得对应的最佳匹配序列。

## 融合深度图和三维模型的人体运动捕捉方法

### 技术领域

[0001] 本发明属于多媒体信息检索技术领域,具体涉及一种融合深度图和三维模型的人体运动捕捉方法。

### 背景技术

[0002] 人体运动捕获技术是多媒体信息检索领域的热点问题,特别是在影视动画、游戏等的发展中,具有广泛应用前景,国内外许多研究机构正致力于这一方向的研究。近年来,随着运动捕获技术的快速发展,三维影视动画、游戏、新一代人交互等领域的兴起,非常多复杂和逼真的人体动作需要快速的被捕获应用,需要一种快速有效的方法进行人体运动的捕获。目前已提出的光学式的运动捕获方法,主要基于计算机视觉原理,通过对目标上特定光点的监视和跟踪来完成运动捕捉任务。但上述的运动捕获方法有几点不足:

[0003] (1) 光学式的运动捕获在使用时需要在表演者身上粘贴标记点,并需要表演者穿上特制的表演用服装,交互起来并不方便。当运动复杂时,不同部位的标记点有可能发生混淆、遮挡,产生错误结果,这时需要人工干预后处理过程。

[0004] (2) 虽然它可以捕捉实时运动,但后处理(包括标记点的识别、跟踪、空间坐标的计算)的工作量较大,对于表演场地的光照、反射情况有一定的要求,装置定标也较为烦琐。

### 发明内容

[0005] 本发明的目的是提供一种融合深度图和三维模型的人体运动捕捉方法,能有效克服现有运动捕获方法中动作幅度受限制、捕捉效果失真度高和误差较大的技术缺陷。

[0006] 本发明所采用的技术方案是:

[0007] 融合深度图和三维模型的人体运动捕捉方法,其特征在于:

[0008] 由以下步骤实现:

[0009] 步骤一:采集人体动作的深度信息,去除运动目标背景,获取完整的人体动作深度信息;

[0010] 步骤二:将提取的人体动作深度信息信息转化成人体的三维点云信息,对人体进行三维重建,获得人体动作的三维模型;

[0011] 步骤三:重复步骤一和步骤二,基于大量采集的人体动作深度信息,建立人体动作的三维模型数据库 $M = \{Y_1, Y_2, \dots, Y_n\}$ ,  $M$ 表示三维模型数据库,  $Y_n$ 表示第 $n$ 个三维模型;

[0012] 步骤四:根据人体结构及人体动作的三维模型构建出人体动作骨架,建立人体动作的骨架数据库 $G = \{S_1, S_2, \dots, S_n\}$ ,  $G$ 表示人体骨架数据库,其中 $S_n$ 表示第 $n$ 个人体骨架数据,其中人体动作的骨架数据与三维模型数据一一对应;

[0013] 步骤五:提取待识别人体动作的深度信息,基于深度信息构建待识别人体动作的三维模型,然后与三维模型数据库中的人体动作进行相似性匹配,按相似性排序输出人体动作骨架,相似性距离值最小的骨架作为最优骨架,以最优骨架作为运动捕获结果;

[0014] 步骤一中,基于采集的人体动作深度信息,去除运动目标背景,获取完整的人体动

作深度信息的具体步骤为:

[0015] (1) 采集的一个人体动作深度图用 $F(x, y, d_p)$ 表示, 其中, $x$ 、 $y$ 分别为像素坐标系下的横坐标和纵坐标, $d_p$ 为深度信息;假设基于深度信息分割背景区域和目标区域的阈值为:

[0016]  $T = (\maxDepthVulue + \minDepthVulue) / 2$

[0017] 其中, $\maxDepthVulue$ 和 $\minDepthVulue$ 分别为图像深度值的最大和最小值,将 $T$ 记录在 $T_0$ 中,根据阈值 $T$ ,将 $F(x, y, d_p)$ 分为背景区域和目标区域,分别求出背景区域和目标区域的平均深度值 $du_1$ 和 $du_2$ ;

[0018] (2) 重新计算 $T = (du_1 + du_2) / 2$ ,判断 $T$ 与 $T_0$ 是否相等,如果不相等,那么将 $T$ 记录在 $T_0$ 里,重复上述步骤,直到 $T = T_0$ 成立,终止算法;把最后得到的 $T$ 作为最佳阈值对 $F(x, y, d_p)$ 进行分割,去除背景,得到完整的人体动作深度信息 $d_0$ ;

[0019] 步骤二中,将提取的人体动作深度信息转化成人体的三维点云信息,对人体进行三维重建,获得人体动作的三维模型的具体步骤为:

[0020] (1) 将提取出的深度信息 $d_0$ 归一化,假设深度值的个数为 $N$ ,计算出深度值中的最大值 $\max d_0(k)$ 和最小值 $\min d_0(k)$ ,归一化后的深度值为:

[0021]  $z(n) = (d_0(n) - \min d_0(k)) / (\max d_0(k) - \min d_0(k))$ ,  $n = 1, 2, \dots, N$

[0022] (2) 令 $Z = z(n)$ ,世界坐标系是以摄像机为原点的,经过标定后摄像机可以看做是理想的成像模型,根据简单的相似三角形变换原理:

$$[0023] \quad \frac{X}{x} = Z \times \frac{1}{f}$$

$$[0024] \quad \frac{Y}{y} = Z \times \frac{1}{f}$$

[0025] 从中可以计算得到世界坐标系的 $X$ 、 $Y$ 值,其中, $x$ 、 $y$ 分别为像素坐标系的横坐标和纵坐标, $f$ 为摄像机的焦距;最终可以得到人体动作的三维点云信息 $(X, Y, Z)$ ;

[0026] 步骤四中,根据人体结构及人体动作的三维模型构建出人体动作骨架的具体步骤为:

[0027] (1) 在深度图中将身体躯干近似视为四边形,记为 $Q$ ;

[0028]  $Q$ 上方的两个顶点就是肩关节点 $a_1$ 和 $a_2$ 的位置;

[0029] 颈关节点 $b$ 为 $a_1$ 、 $a_2$ 连线的中点;

[0030] 沿 $b$ 点向上,最上端就是头部节点 $c$ 的位置;

[0031]  $Q$ 下方的两个顶点为髋关节 $d_1$ 和 $d_2$ 的位置;

[0032] 胯关节点 $e$ 在 $d_1$ 、 $d_2$ 的中点位置;

[0033] (2) 确定手关节点和肘关节位置从 $a_1$ 、 $a_2$ 处开始进行搜索,若手臂伸直,根据上下手臂长度比例确定肘关节点 $f_1$ 、 $f_2$ 和手关节点 $g_1$ 、 $g_2$ 的位置;若手臂弯曲,在肘关节点的位置有拐点,拐点的位置为 $f_1$ 、 $f_2$ ,沿 $f_1$ 、 $f_2$ 继续搜索至终点为手关节点 $g_1$ 、 $g_2$ 的位置;

[0034] 同理,根据手关节点和肘关节位置的确定方法可以确定出膝关节点 $h_1$ 、 $h_2$ 和脚关节点 $i_1$ 、 $i_2$ 的位置;

[0035] (3) 将各个关节点坐标按人体结构顺序用直线相连接,就得到了人体动作骨架;

[0036] 步骤五中,提取待识别人体动作的深度信息,基于深度信息构建待识别人体动作

的三维模型,然后与三维模型数据库中的人体动作进行相似性匹配的具体步骤为:

[0037] (1) 将三维模型表示的人体动作层次聚类算法进行聚类:

[0038] ①假设一个参数 $m$ 表示最终聚类簇的数目,将每个输入的数据作为单独的一个数据簇 $D_0$ ,求出与每个 $D$ 相邻的最近的簇 $D_x$ ,数据簇间的距离可以看做是数据簇中心间的距离;

[0039] ②将距离最近的两个数据簇 $D_p$ 和 $D_q$ 进行合并,产生的新的数据簇 $D_n$ ,再计算 $D_n$ 与其它簇的距离,如果当前数据簇的数目大于 $m$ ,那么就转到②继续进行数据簇的合并,否则算法结束,得到聚类结果 $D = \{u_1, u_2, \dots, u_m\}$ ,其中 $u$ 表示每个类的值;

[0040] (2) 计算待识别人体动作序列各个动作帧与三维模型数据库中的人体动作序列各个动作帧之间的距离:

[0041] ①假设待识别的人体动作和数据库中任意一个三维模型聚类结果分别表示为:

[0042]  $D_i = \{u'_1, u'_2, \dots, u'_m\}$

[0043]  $D_j = \{v_1, v_2, \dots, v_m\}$

[0044] 从聚类树的根节点开始,进行深度优先搜索直至叶节点;

[0045] ②在遍历搜索的过程中对 $D_i$ 和 $D_j$ 中各个类的对象计算距离,将各个类的距离求和,得到两颗聚类树的欧式距离:

$$[0046] \quad dist = \sum_{\substack{i=1 \dots m \\ j=1 \dots m}} \sqrt{(u'_{ix} - v_{jx})^2 + (u'_{iy} - v_{jy})^2 + (u'_{iz} - v_{jz})^2}$$

[0047] (3) 用正则时间规整的方法计算出最优匹配路径,得到最优匹配序列:

[0048] ①假设待识别的三维模型动作序列为 $\mathbf{X} = [x_1, \dots, x_m] \in R^{d_x \times m}$ ,三维模型数据库中一组动作序列为 $\mathbf{Y} = [y_1, \dots, y_n] \in R^{d_y \times n}$ , $m$ 和 $n$ 分别表示两个动作序列的长度,两个序列中各个动作帧之间的距离用上述方法可以计算得到;

[0049] ②因为 $m$ 和 $n$ 可能不相等,所以其对应的匹配关系可能有很多种,利用正则时间规整的方法,计算出序列间的最小距离 $J$ ,同时获得最佳匹配路径 $P_x$ 和 $P_y$ :

[0050]

$$\{\mathbf{V}_x, \mathbf{V}_y\} \in \Phi, \{\mathbf{p}_x, \mathbf{p}_y\} \in \Psi \quad J_{ctw} = \left\| \mathbf{V}_x^T \mathbf{X} \mathbf{W}_x - \mathbf{V}_y^T \mathbf{Y} \mathbf{W}_y \right\|_F^2 + \phi(\mathbf{V}_x) + \phi(\mathbf{V}_y)$$

[0051] 这里, $\mathbf{W}_x = \mathbf{W}(\mathbf{p}_x) \in \{0, 1\}^{m \times 1}$ 和 $\mathbf{W}_y = \mathbf{W}(\mathbf{p}_y) \in \{0, 1\}^{n \times 1}$ 是两个二进制的规整矩阵,对每一步的 $t \in \{1:1\}$ 有 $w_{p_t, t} = 1$ ,其他情况时 $w_{p_t, t} = 0$ ,其中1是正则时间规整算法自动选择的需要匹配的动作帧的数目, $1 \geq \max(m, n)$ ;  $\phi(\cdot)$ 是正则项:

$$[0052] \quad \phi(\mathbf{V}) = \frac{\lambda}{1 - \lambda} \left\| \mathbf{V} \right\|_F^2$$

[0053]  $P_x$ 和 $P_y$ 表示所有可能的匹配路径,需要满足的约束条件为:

[0054]  $\Psi = \{ \{P_x, P_y\} \mid P_x \in (1:m)^1, P_y \in (1:n)^1 \}$

[0055] 边界: $[p_1^x, p_1^y] = [1, 1], [p_l^x, p_l^y] = [m, n]$

[0056] 单调性: $t_1 \geq t_2 \Rightarrow p_{t_1}^x \geq p_{t_2}^x, p_{t_1}^y \geq p_{t_2}^y$



[0057] 连续性:  $[p_t^x, p_t^y] - [p_{t-1}^x, p_{t-1}^y] \in \{[0,1], [1,0], [1,1]\}$

[0058]  $V_x \in R^{d_x \times d}$  和  $V_y \in R^{d_y \times d}$  是线性变化矩阵,  $d \leq \min(d_x, d_y)$ , 需满足的约束条件为:

[0059]  $\Phi = \{ \{V_x, V_y\} V_x^T ((1-\lambda) X W_x W_x^T X^T + \lambda I) V_x = I,$

[0060]  $V_y^T ((1-\lambda) Y W_y W_y^T Y^T + \lambda I) V_y = I \}$

[0061] 其中,  $\lambda \in [0, 1]$ ;

[0062] ③重复以上步骤, 利用正则时间规整方法, 计算三维模型数据库中每一组动作序列Y分别与X的最小距离J及最佳匹配路径 $P_x$ 和 $P_y$ ; 然后求出各个J的最小值和其对应的 $P_x$ 和 $P_y$ , 从而获得对应的最佳匹配序列。

[0063] 本发明具有以下优点:

[0064] (1) 本发明所涉及的方法使用相关设备采集深度信息, 表演者无需在人体上安装传感器或添加标记点, 活动的幅度大, 对于表演场地要求不多, 方便且容易实现。

[0065] (2) 本发明针对以往无标记点的运动捕获技术主要是基于视频方法, 拍摄出二维图像序列达到运动捕获的目的, 本发明融合了深度信息, 解决了在没有深度信息的二维图像中身体的自遮挡导致身体部位信息缺失的问题, 保证了信息获取的完整性和可靠性。

[0066] (3) 本发明使用层次聚类法, 数据簇间距离和规则的相似度容易定义, 将数据量较大的三维点云数据缩小为几个类, 以便于后续利用深度优先搜索方法计算两帧三维模型的欧式距离, 减少了计算量。

[0067] (4) 本发明在运动序列匹配上使用正则时间规整(Canonical Time Warping, CTW)的方法, 该方法具有算法鲁棒的优点, 并且即使测试序列模式与数据库中参考序列模式的时间尺度不能完全一致, 它仍能较好地完成测试序列与参考序列之间的模式匹配。CTW是基于动态规划的方法, 步骤五第(3)步的②中公式是CTW方法的目标函数, 目标函数中加有正则项和两个二进制规整矩阵, 为目标函数提供了唯一解的可能性, 提高了两个序列匹配的精度, 使匹配的时间大幅降, 保证了运动捕获的速度和精度。

## 附图说明

[0068] 图1为本发明方法的总体流程图。

[0069] 图2为本发明步骤一的具体流程框图。

[0070] 图3为本发明步骤五的具体流程框图。

## 具体实施方式

[0071] 下面结合具体实施方式对本发明进行详细的说明。

[0072] 本发明所涉及的融合深度图和三维模型的人体运动捕捉方法, 由以下步骤实现:

[0073] 步骤一: 应用相关设备采集人体动作的深度信息, 去除运动目标背景, 获取完整的人体动作深度信息。



[0074] 基于采集的人体动作深度信息,去除运动目标背景,获取完整的人体动作深度信息的具体步骤为:

[0075] (1) 采集的一个人体动作深度图用 $F(x, y, d_p)$ 表示,其中, $x$ 、 $y$ 分别为像素坐标系下的横坐标和纵坐标, $d_p$ 为深度信息;假设基于深度信息分割背景区域和目标区域的阈值为:

[0076]  $T = (\maxDepthVulue + \minDepthVulue) / 2$

[0077] 其中, $\maxDepthVulue$ 和 $\minDepthVulue$ 分别为图像深度值的最大和最小值,将 $T$ 记录在 $T_0$ 中,根据阈值 $T$ ,将 $F(x, y, d_p)$ 分为背景区域和目标区域,分别求出背景区域和目标区域的平均深度值 $du_1$ 和 $du_2$ ;

[0078] (2) 重新计算 $T = (du_1 + du_2) / 2$ ,判断 $T$ 与 $T_0$ 是否相等,如果不相等,那么将 $T$ 记录在 $T_0$ 里,重复上述步骤,直到 $T = T_0$ 成立,终止算法;把最后得到的 $T$ 作为最佳阈值对 $F(x, y, d_p)$ 进行分割,去除背景,得到完整的人体动作深度信息 $d_0$ 。

[0079] 步骤二:将提取的人体动作深度信息转化成人体的三维点云信息,对人体进行三维重建,获得人体动作的三维模型,具体步骤为:

[0080] (1) 将提取出的深度信息 $d_0$ 归一化,假设深度值的个数为 $N$ ,计算出深度值中的最大值 $\max d_0(k)$ 和最小值 $\min d_0(k)$ ,归一化后的深度值为:

[0081]  $z(n) = (d_0(n) - \min d_0(k)) / (\max d_0(k) - \min d_0(k))$ ,  $n = 1, 2, \dots, N$

[0082] (2) 令 $Z = z(n)$ ,世界坐标系是以摄像机为原点的,经过标定后摄像机可以看做是理想的成像模型,根据简单的相似三角形变换原理:

[0083]  $\frac{X}{x} = Z \times \frac{1}{f}$

[0084]  $\frac{Y}{y} = Z \times \frac{1}{f}$

[0085] 从中可以计算得到世界坐标系的 $X$ 、 $Y$ 值,其中, $x$ 、 $y$ 分别为像素坐标系的横坐标和纵坐标, $f$ 为摄像机的焦距。最终可以得到人体动作的三维点云信息 $(X, Y, Z)$ 。

[0086] 步骤三:重复步骤一和步骤二,基于大量采集的尽可能多的人体动作深度信息,建立人体动作的三维模型数据库 $M = \{Y_1, Y_2, \dots, Y_n\}$ , $M$ 表示三维模型数据库, $Y_n$ 表示第 $n$ 个三维模型。

[0087] 步骤四:根据人体结构及人体动作的三维模型构建出人体动作骨架,建立人体动作的骨架数据库 $G = \{S_1, S_2, \dots, S_n\}$ , $G$ 表示人体骨架数据库,其中 $S_n$ 表示第 $n$ 个人体骨架数据,其中人体动作的骨架数据与三维模型数据一一对应。

[0088] 根据人体结构及人体动作的三维模型构建出人体动作骨架的具体步骤为:

[0089] (1) 在深度图中将身体躯干近似视为四边形,记为 $Q$ ;

[0090]  $Q$ 上方的两个顶点就是肩关节点 $a_1$ 和 $a_2$ 的位置;

[0091] 颈关节点 $b$ 为 $a_1$ 、 $a_2$ 连线的中点;

[0092] 沿 $b$ 点向上,最上端就是头部节点 $c$ 的位置;

[0093]  $Q$ 下方的两个顶点为髋关节 $d_1$ 和 $d_2$ 的位置;

[0094] 胯关节点 $e$ 在 $d_1$ 、 $d_2$ 的中点位置;

[0095] (2) 确定手关节点和肘关节位置从 $a_1$ 、 $a_2$ 处开始进行搜索,若手臂伸直,根据上下手

臂长度比例确定肘关节点 $f_1$ 、 $f_2$ 和手关节点 $g_1$ 、 $g_2$ 的位置;若手臂弯曲,在肘关节点的位置有拐点,拐点的位置为 $f_1$ 、 $f_2$ ,沿 $f_1$ 、 $f_2$ 继续搜索至终点为手关节点 $g_1$ 、 $g_2$ 的位置;

[0096] 同理,根据手关节点和肘关节位置的确定方法可以确定出膝关节点 $h_1$ 、 $h_2$ 和脚关节点 $i_1$ 、 $i_2$ 的位置;

[0097] (3) 将各个关节点坐标按人体结构顺序用直线相连接,就得到了人体动作骨架。

[0098] 步骤五:提取待识别人体动作的深度信息,基于深度信息构建待识别人体动作的三维模型,然后与三维模型数据库中的人体动作进行相似性匹配,按相似性排序输出人体动作骨架,相似性距离值最小的骨架作为最优骨架,以最优骨架作为运动捕获结果。

[0099] 提取待识别人体动作的深度信息,基于深度信息构建待识别人体动作的三维模型,然后与三维模型数据库中的人体动作进行相似性匹配的具体步骤为:

[0100] (1) 将三维模型表示的人体动作层次聚类算法 (THA) 进行聚类:

[0101] ①假设一个参数 $m$ 表示最终聚类簇的数目,将每个输入的数据作为单独的一个数据簇 $D_0$ ,求出与每个 $D$ 相邻的最近的簇 $D_x$ ,数据簇间的距离可以看做是数据簇中心间的距离;

[0102] ②将距离最近的两个数据簇 $D_p$ 和 $D_q$ 进行合并,产生的新的数据簇 $D_n$ ,再计算 $D_n$ 与其它簇的距离,如果当前数据簇的数目大于 $m$ ,那么就转到②继续进行数据簇的合并,否则算法结束,得到聚类结果 $D = \{u_1, u_2, \dots, u_m\}$ ,其中 $u$ 表示每个类的值;

[0103] (2) 计算待识别人体动作序列各个动作帧与三维模型数据库中的人体动作序列各个动作帧之间的距离:

[0104] ①假设待识别的人体动作和数据库中任意一个三维模型聚类结果分别表示为:

[0105]  $D_i = \{u'_1, u'_2, \dots, u'_m\}$

[0106]  $D_j = \{v_1, v_2, \dots, v_m\}$

[0107] 从聚类树的根节点开始,进行深度优先搜索直至叶节点;

[0108] ②在遍历搜索的过程中对 $D_i$ 和 $D_j$ 中各个类的对象计算距离,将各个类的距离求和,得到两颗聚类树的欧式距离:

$$[0109] \quad dist = \sum_{\substack{i=1 \dots m \\ j=1 \dots m}} \sqrt{(u'_{ix} - v_{jx})^2 + (u'_{iy} - v_{jy})^2 + (u'_{iz} - v_{jz})^2}$$

[0110] (3) 用正则时间规整 (Canonical Time Warping, CTW) 的方法计算出最优匹配路径,得到最优匹配序列:

[0111] ①假设待识别的三维模型动作序列为 $\mathbf{X} = [x_1, \dots, x_m] \in R^{d_x \times m}$ ,三维模型数据库中的一组动作序列为 $\mathbf{Y} = [y_1, \dots, y_n] \in R^{d_y \times n}$ , $m$ 和 $n$ 分别表示两个动作序列的长度,两个序列中各个动作帧之间的距离用上述方法可以计算得到;

[0112] ②因为 $m$ 和 $n$ 可能不相等,所以其对应的匹配关系可能有很多种,利用正则时间规整的方法,计算出序列间的最小距离 $J$ ,同时获得最佳匹配路径 $P_x$ 和 $P_y$ :

[0113]

$$\min_{\{\mathbf{V}_x, \mathbf{V}_y\} \in \Phi, \{\mathbf{p}_x, \mathbf{p}_y\} \in \Psi} J_{ctw} = \left\| \mathbf{V}_x^T \mathbf{X} \mathbf{W}_x - \mathbf{V}_y^T \mathbf{Y} \mathbf{W}_y \right\|_F^2 + \phi(\mathbf{V}_x) + \phi(\mathbf{V}_y)$$

[0114] 这里,  $W_x = W(p_x) \in \{0, 1\}^{m \times 1}$  和  $W_y = W(p_y) \in \{0, 1\}^{n \times 1}$  是两个二进制的规整矩阵, 对每一步的  $t \in \{1:1\}$  有  $w_{p_t, t} = 1$ , 其他情况时  $w_{p_t, t} = 0$ , 其中 1 是正则时间规整算法自动选择的需要匹配的动作帧的数目,  $1 \geq \max(m, n)$ 。  $\Phi(\cdot)$  是正则项:

$$[0115] \quad \phi(V) = \frac{\lambda}{1-\lambda} \|V\|_F^2$$

[0116]  $P_x$  和  $P_y$  表示所有可能的匹配路径, 需要满足的约束条件为:

$$[0117] \quad \Psi = \{ \{P_x, P_y\} \mid P_x \in (1:m)^1, P_y \in (1:n)^1 \}$$

$$[0118] \quad \text{边界: } [p_1^x, p_1^y] = [1, 1], [p_l^x, p_l^y] = [m, n]$$

$$[0119] \quad \text{单调性: } t_1 \geq t_2 \Rightarrow p_{t_1}^x \geq p_{t_2}^x, p_{t_1}^y \geq p_{t_2}^y$$

$$[0120] \quad \text{连续性: } [p_t^x, p_t^y] - [p_{t-1}^x, p_{t-1}^y] \in \{[0, 1], [1, 0], [1, 1]\}$$

[0121]  $V_x \in R^{d_x \times d}$  和  $V_y \in R^{d_y \times d}$  ( $d \leq \min(d_x, d_y)$ ) 是线性变化矩阵, 需满足的约束条件为:

$$[0122] \quad \Phi = \left\{ \left\{ V_x, V_y \right\} \left\{ V_x^T \left( (1-\lambda) X W_x W_x^T X^T + \lambda I \right) V_x = I, \right. \right.$$

$$[0123] \quad \left. V_y^T \left( (1-\lambda) Y W_y W_y^T Y^T + \lambda I \right) V_y = I \right\}$$

[0124] 其中,  $\lambda \in [0, 1]$ ;

[0125] ③重复以上步骤, 利用正则时间规整方法, 计算三维模型数据库中每一组动作序列  $Y$  分别与  $X$  的最小距离  $J$  及最佳匹配路径  $P_x$  和  $P_y$ 。然后求出各个  $J$  的最小值和其对应的  $P_x$  和  $P_y$ , 从而获得对应的最佳匹配序列。

[0126] 本发明的内容不限于实施例所列举, 本领域普通技术人员通过阅读本发明说明书而对本发明技术方案采取的任何等效的变换, 均为本发明的权利要求所涵盖。

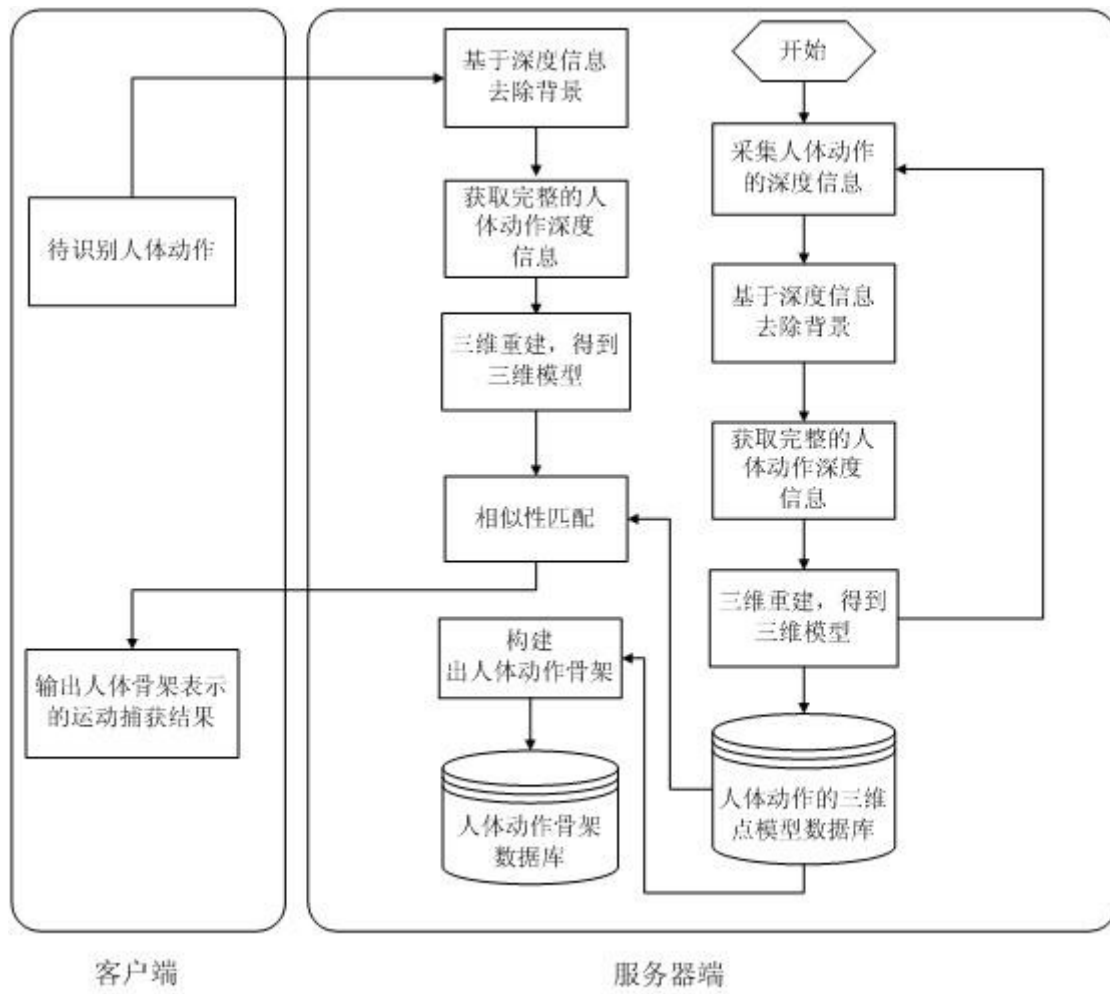


图1

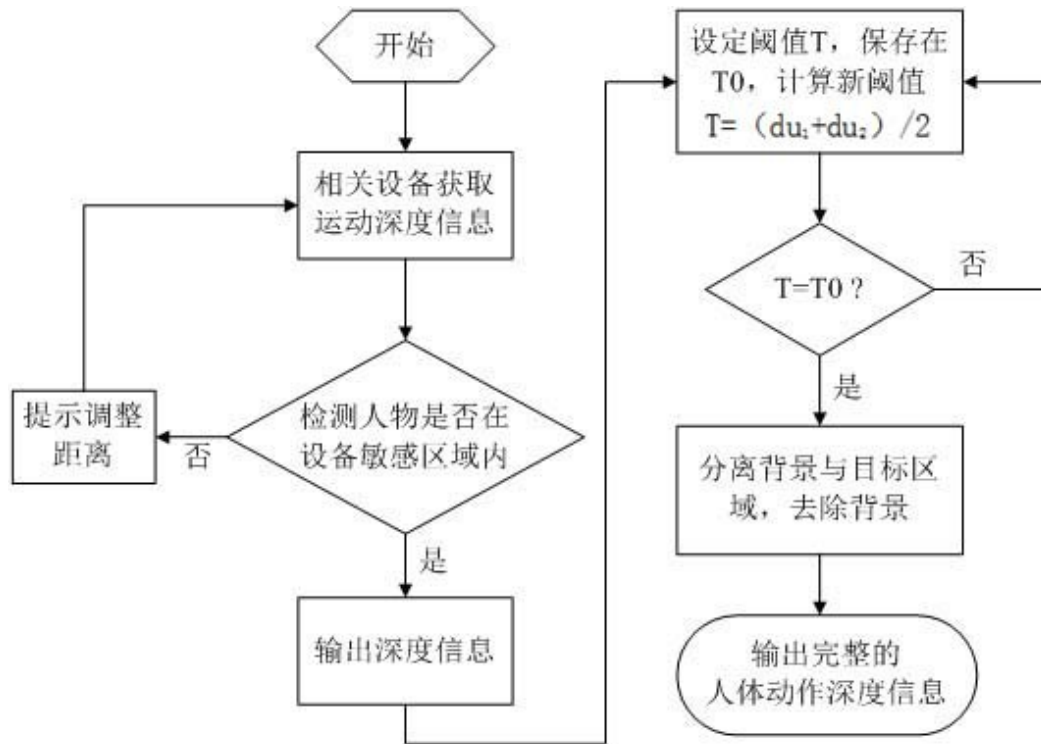


图2

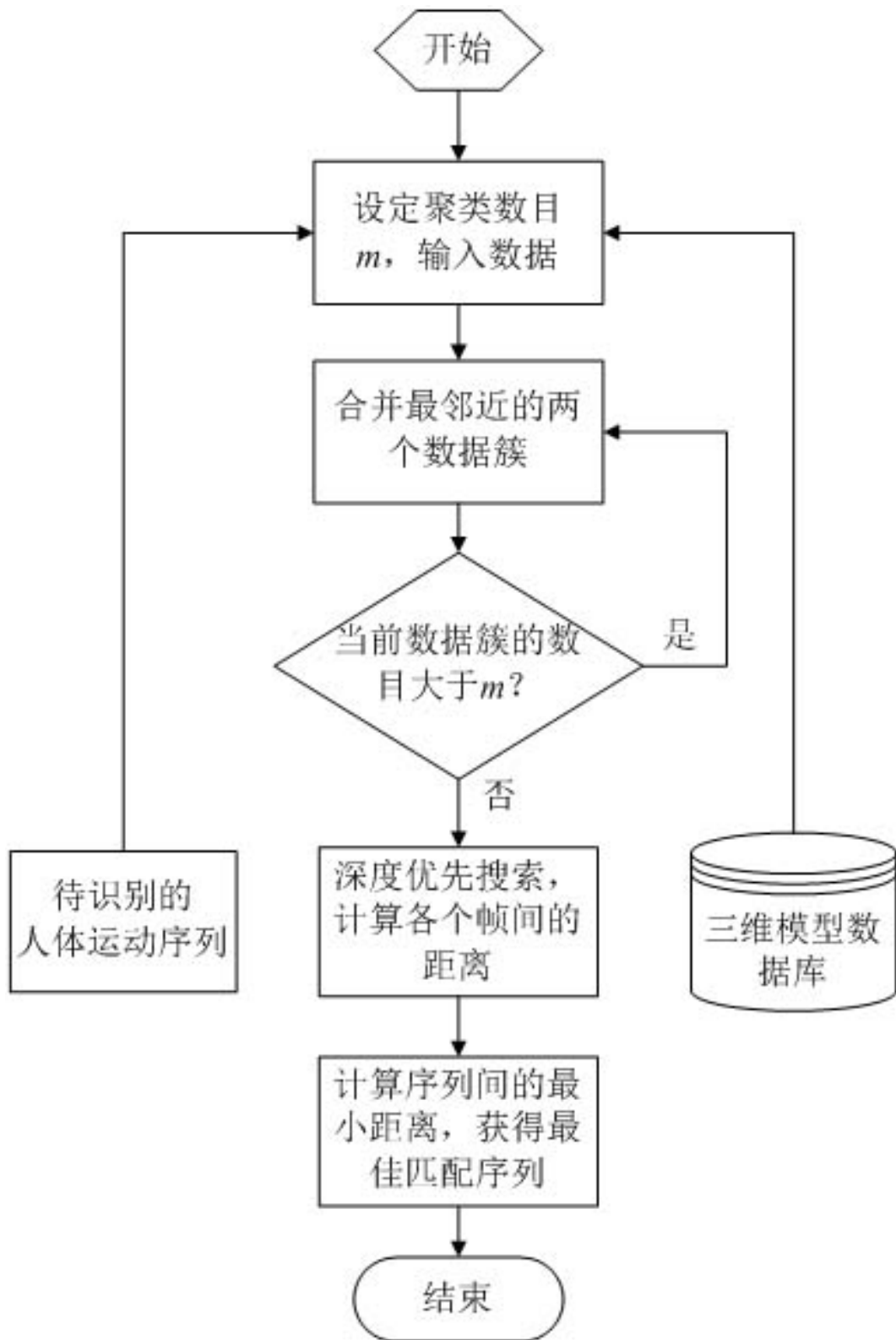


图3