



(12)发明专利

(10)授权公告号 CN 106648103 B

(45)授权公告日 2019.09.27

(21)申请号 201611236595.6

(22)申请日 2016.12.28

(65)同一申请的已公布的文献号
申请公布号 CN 106648103 A

(43)申请公布日 2017.05.10

(73)专利权人 歌尔科技有限公司
地址 266104 山东省青岛市崂山区北宅街道投资服务中心308室

(72)发明人 仇璐 张绍谦 张超

(74)专利代理机构 青岛联智专利商标事务所有
限公司 37101

代理人 周容

(51)Int.Cl.

G06F 3/01(2006.01)

G06K 9/00(2006.01)

(56)对比文件

CN 101839701 A,2010.09.22,

CN 104112289 A,2014.10.22,

US 2014044344 A1,2014.02.13,

于晓磊.基于法向量和曲率结合的深度图像分割.《中国优秀硕士学位论文全文数据库信息科技辑》.2012,(第5期),全文.

审查员 宫玉龙

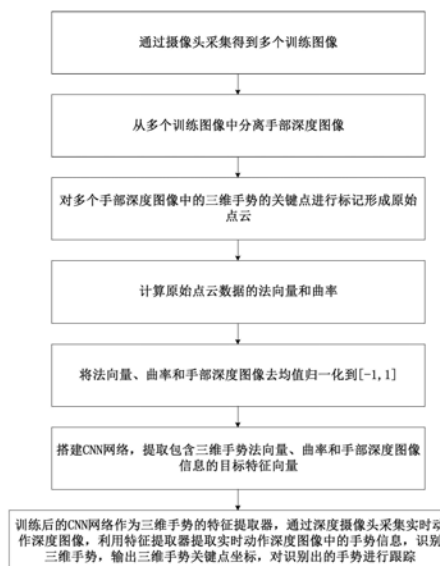
权利要求书2页 说明书5页 附图2页

(54)发明名称

一种VR头戴设备的手势跟踪方法和VR头戴设备

(57)摘要

本发明提供一种VR头戴设备的手势跟踪方法,包括以下步骤:采集多个训练图像;分离手部深度图像;标记三维手势,形成原始点云;计算法向量和曲率,去均值归一化;搭建CNN网络,CNN网络的输入端分别输入多法向量、曲率和手部深度图像,输出端输出包括掌心在内的多个关节的三维坐标;训练后的CNN网络作为三维手势的特征提取器,通过深度摄像头采集实时动作深度图像,特征提取器提取并处理实时动作深度图像中所包含的三维手势的法向量、曲率和手部深度图像信息,输出包含掌心的多个关节的三维坐标,并对识别出的三维手势进行跟踪。还公开一种VR头戴设备。本发明融合了三维特征信息,具有模型识别率高的优点。



1. 一种VR头戴设备的手势跟踪方法,其特征在于,包括以下步骤:

通过摄像头采集得到多个训练图像;

从多个所述训练图像中分离多个手部深度图像;

对多个所述手部深度图像中的三维手势的关键点进行标记,利用手部深度图像形成原始点云数据,所述关键点包括掌心和多个手部关节;

计算根据手部深度图像形成的原始点云数据的法向量和曲率;

将法向量、曲率和手部深度图像去均值归一化到 $[-1, 1]$;

搭建CNN网络,所述CNN网络的输入端分别输入根据多个训练图像生成并归一化后的法向量、曲率和手部深度图像三路数据,输出端输出包括掌心在内的多个关节的三维坐标;

训练完成的CNN网络作为三维手势的特征提取器,通过深度摄像头采集实时动作深度图像,所述特征提取器提取所述实时动作深度图像中的法向量、曲率和手部深度图像信息,输出实时动作深度图像中三维手势的三维坐标,处理器对识别出的三维手势进行跟踪。

2. 根据权利要求1所述的VR头戴设备的手势跟踪方法,其特征在于:

计算所述原始点云数据的法向量和曲率时,使用kd-tree算法构建所述原始点云的树结构,并利用原始点云的树结构查找计算手部深度图像原始点云数据的法向量和曲率。

3. 根据权利要求2所述的VR头戴设备的手势跟踪方法,其特征在于:从所述训练图像中分离手部深度图像包括以下步骤:

利用随机森林算法将通过摄像头采集得到的多幅训练图像中的手部深度图像与背景深度数据分离;

对分离出的所述手部深度图像进行降噪。

4. 根据权利要求3所述的VR头戴设备的手势跟踪方法,其特征在于,还包括以下步骤:

将降噪后的手部深度图像归一化降维,生成 256×256 的二维图像;

将降噪后的手部深度图像中的关键点进行标记,所述关键点包括多个关节和掌心;

生成所述手部深度图像中任一点的图像坐标 $m = (x, y)^T$;

利用所述手部深度图像中任一点的图像坐标 $m = (x, y)^T$ 和相机坐标系下的空间坐标的关系求解所述手部深度图像中任一点的相机坐标系中的空间坐标 x_c, y_c 和 z_c ,图像坐标和相机坐标系下空间坐标的关系如下:

$$z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, \text{其中 } z_c \text{ 为相机的光轴};$$

在相机坐标系下建立对应所述手部深度图像且包括多个关节和掌心坐标的原始点云;

利用kd-tree算法构建对应原始点云数据的树结构,并分别利用原始点云数据的树结构查找计算法向量和曲率。

5. 根据权利要求4所述的VR头戴设备的手势跟踪方法,其特征在于:

搭建CNN网络之前,利用PCA算法对手部深度图像降维到 96×96 。

6. 根据权利要求5所述的VR头戴设备的手势跟踪方法,其特征在于:

所述CNN网络包括卷积层、池化层和全连接层,其中所述卷积层包括并行的第一卷积通

道、第二卷积通道和第三卷积通道,所述第一卷积通道、第二卷积通道和第三卷积通道的输入端分别输入法向量、曲率和手部深度图像。

7. 根据权利要求6所述的VR头戴设备的手势跟踪方法,其特征在于:

所述第一卷积通道、第二卷积通道和第三卷积通道均包括三级卷积,且每一级卷积层后均跟随一层池化层,第一卷积通道、第二卷积通道和第三卷积通道输出至全连接层,所述全连接层包括三级全连接层。

8. 根据权利要求7所述的VR头戴设备的手势跟踪方法,其特征在于,所述CNN网络的激活函数为Relu函数。

9. 根据权利要求8所述的VR头戴设备的手势跟踪方法,其特征在于,所述卷积层的卷积核为 5×5 ,所述池化层的池化核为 2×2 。

10. 一种VR头戴设备,其特征在于,采用如权利要求1至9任一项所示的VR头戴设备的手势跟踪方法。

一种VR头戴设备的手势跟踪方法和VR头戴设备

技术领域

[0001] 本发明涉及虚拟现实技术领域,尤其涉及一种VR头戴设备的手势跟踪方法以及VR头戴设备。

背景技术

[0002] 虚拟现实技术是仿真技术的一个重要分支方向。虚拟现实技术使用计算机,利用相关技术和软、硬件工具生成实时动态的、三维立体而且纹理逼真的图像和场景,并使其可以模仿人类的各种感知,并利用传感器与用户进行交互。从1963年虚拟现实技术萌芽至今,虚拟现实技术的理论已经比较完善,而且近年来,虚拟现实技术在军事仿真、娱乐游戏、医疗、建筑等多个行业中得到广泛和深入的研究和使用。

[0003] 在现有的虚拟现实设备的人机交互过程中,除了传统的按键操作之外,还有手势识别功能,VR头戴设备通过设置在设备中的摄像头采集视野中的图像,并从图像中分离识别手部图像,对手部图像进行模型匹配来判别手势类别或者跟踪手势坐标。其中分离识别动作建立模型训练的基础上且通常采用卷积神经网络进行手势数据训练。

[0004] 在现有的虚拟现实设备跟踪手势坐标的过程中,通常是采集手势深度数据并利用CNN进行回归训练得到模型。而现有CNN网络的训练核心是对二维图卷积提取特征,仅仅利用手势深度数据进行训练,所获得的信息可以说是从二维平面上提取的信息,手部的三维空间立体结构基本没有得到有效利用,由于CNN采集到的信息是平面信息,所以CNN的训练难度大,得到的数据误差大、跟踪的手势坐标也不够准确。

发明内容

[0005] 为解决现有VR头戴设备图像采集时,手部的三维空间立体结构基本没有得到有效利用,导致数据误差大、跟踪手势坐标不准确的问题,本发明公开了一种VR头戴设备的手势跟踪方法。

[0006] 一种VR头戴设备的手势跟踪方法,包括以下步骤:

[0007] 通过摄像头采集得到多个训练图像;

[0008] 从多个所述训练图像中分离多个手部深度图像;

[0009] 对多个所述手部深度图像中的三维手势的关键点进行标记,利用手部深度图像形成原始点云数据,所述关键点包括掌心和多个手部关节;

[0010] 计算根据手部深度图像形成的原始点云数据的法向量和曲率;

[0011] 将法向量、曲率和手部深度图像去均值归一化到 $[-1, 1]$;

[0012] 搭建CNN网络,所述CNN网络的输入端分别输入根据多个训练图像生成并归一化后的法向量、曲率和手部深度图像三路数据,输出端输出包括掌心在内的多个关节点的三维坐标;

[0013] 训练完成的CNN网络作为三维手势的特征提取器,通过深度摄像头采集实时动作深度图像,所述特征提取器提取所述实时动作深度图像中的法向量、曲率和手部深度图像

信息,输出实时动作深度图像中三维手势的三维坐标,所述处理器对识别出的三维手势进行跟踪。

[0014] 进一步的,计算所述原始点云数据的法向量和曲率时,使用kd-tree算法构建所述原始点云的树结构,并利用原始点云的树结构查找计算手部深度图像原始点云数据的法向量和曲率。

[0015] 进一步的,利用随机森林算法将通过摄像头采集得到的多幅训练图像中的手部深度图像与背景深度数据分离;对所述手部深度图像进行降噪。

[0016] 进一步的,将降噪后的手部深度图像归一化降维,生成 256×256 的二维图像;

[0017] 将降噪后的手部深度图像中的关键点进行标记,所述关键点包括多个关节点和掌心;

[0018] 生成手部深度图像中任一点的图像坐标 $m = (x, y)^T$;

[0019] 利用手部深度图像中任一点的图像坐标 $m = (x, y)^T$ 和相机坐标系下的空间坐标的关系求解所述手部深度图像中任一点的相机坐标系中的空间坐标 x_c, y_c 和 z_c ,图像坐标和相机坐标系下空间坐标的关系如下:

$$[0020] \quad z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, \text{其中 } z_c \text{ 为相机的光轴};$$

[0021] 在相机坐标系下建立对应所述手部深度图像且包括多个关节点和掌心坐标的原始点云;利用kd-tree算法构建对应原始点云数据的树结构,并分别利用原始点云数据的树结构查找计算法向量和曲率。

[0022] 为了提高处理效率,并尽可能的保存图像特征,搭建CNN网络之前,利用PCA算法对手部深度图像降维到 96×96 。

[0023] 进一步的,所述CNN网络包括卷积层、池化层和全连接层,其中所述卷积层包括并行的第一卷积通道、第二卷积通道和第三卷积通道,所述第一卷积通道、第二卷积通道和第三卷积通道的输入端分别输入法向量、曲率和手部深度图像。

[0024] 进一步的,所述第一卷积通道、第二卷积通道和第三卷积通道均包括三级卷积,且每一级卷积层后均跟随一层池化层,第一卷积通道、第二卷积通道和第三卷积通道输出至全连接层,所述全连接层包括三级全连接层。

[0025] 优选的,所述CNN网络的激活函数为Relu函数。

[0026] 优选的,所述卷积层的卷积核为 5×5 ,所述池化层的池化核为 2×2 。

[0027] 本发明在对原始手势深度数据的预处理过程中对数据的法向量、曲率这两种三维空间特征信息进行了提取,并在横向尺度和纵向层深上做了改进。通过对采集数据三维空间描述信息的提取和利用,弥补了手部深度图像特征纹理单一的不足,强化了手势深度数据所具有的三维空间特征,三维手势无需做分类,因此本发明所提供的方法可以适用于对连续变动手势的追踪,同时,由于在采集数据中增加了三维空间的描述信息,因此可以最大可能的排除相机采集角度带来的误差。本发明提高了CNN训练模型的精确度,进而提高了VR头戴设备对手势跟踪的精确度。

[0028] 一种VR头戴设备,采用VR头戴设备的手势跟踪方法控制,所述VR头戴设备的手势

跟踪方法包括以下步骤：

[0029] 通过摄像头采集得到多个训练图像；

[0030] 从多个所述训练图像中分离多个手部深度图像；

[0031] 对多个所述手部深度图像中的三维手势的关键点进行标记，利用手部深度图像形成原始点云数据，所述关键点包括掌心和多个手部关节；

[0032] 计算根据手部深度图像形成的原始点云数据的法向量和曲率；

[0033] 将法向量、曲率和手部深度图像去均值归一化到 $[-1, 1]$ ；

[0034] 搭建CNN网络，所述CNN网络的输入端分别输入根据多个训练图像生成并归一化后的法向量、曲率和手部深度图像三路数据，输出端输出包括掌心在内的多个关节的三维坐标；

[0035] 训练完成的CNN网络作为三维手势的特征提取器，通过深度摄像头采集实时动作深度图像，所述特征提取器提取所述实时动作深度图像中的法向量、曲率和手部深度图像信息，输出实时动作深度图像中三维手势的三维坐标，所述处理器对识别出的三维手势进行跟踪。

[0036] 本发明所公开的VR头戴式设备，所采用的手势跟踪方法中提取融合的三维手势的空间特征，并且通过CNN网络充分利用手部三维信息，提升了卷积神经网络的训练效果，足以满足高精确度的仿真场景，扩大了VR头戴式设备的使用范围。同时由于本发明可以对连续变化的手势实现识别，因此可以脱离实体控制按键，实现远距离多种手部命令控制。

附图说明

[0037] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0038] 图1为本发明所公开的VR头戴设备手势跟踪方法一种实施例的流程图；

[0039] 图2为图1所公开的VR头戴设备手势跟踪方法中所搭建的CNN网络的网络架构图。

具体实施方式

[0040] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0041] 参见图1所示为本发明所公开的VR头戴设备的手势跟踪方法一种实施例的流程示意图。具体来说，本实施例所公开的跟踪方法包括图像预处理、三维特征提取、搭建卷积神经网络、卷积神经网络训练、识别、利用识别结果操控等几个步骤，以下按照上述过程分别详细描述。

[0042] 首先为建立卷积神经网络进行前期准备。通过摄像头采集多个训练样本图像。训练样本图像优选在5000张左右，也可以根据实际识别精度的要求调整训练样本图像的数量，训练样本图像中均包含一个三维手势图像。摄像头可以选择使用设置在VR头戴设备上

的摄像头,也可以选择其它独立设置的高清摄像头。对训练样本图像预处理时,假定三维手势是距离摄像头最近的目标物体,考虑到VR头戴设备的应用场景,对于VR头戴式设备会采集到的图像数据来说,图像背景中人体数据较多,其它因素基本可以忽略。利用随机森林算法将训练图像中的手部深度图像块和背景物体的深度图像分离,提取出对应每一个三维手势图像的手部深度图像块作为手部深度图像,并进一步对手部深度图像进行去噪处理,去除手部深度图像上的噪点,完成图像分离预处理。

[0043] 将降噪后的手部深度图像图像归一化降维,获得分辨率为 256×256 的二维图像。 256×256 的像素选择是基于长期对VR生成图像的处理经验,这一分辨率可以尽可能保持图像信息的完整并降低后续图像处理系统的信息处理量。在本实施例中,识别手势的最终目的是可以自动判定三维手势,并利用三维手势图像特征的变化产生的控制信号进行VR头戴设备的下一步操作。在本方法中,识别的基准在于形成不同手势变化时手部各个指关节会产生对应的变化以及各个指关节会出现不同位置的组合,所以在对手部深度图像预处理时,对手部各个指关节的图像坐标进行标记,同时标记图像中掌心的图像坐标,生成每一个关键点的二维图像坐标。

[0044] 标记完成后,通过计算机系统记录手部深度图像中任一点的图像坐标 $m = (x, y)^T$,利用手部深度图像中任一点的图像坐标 $m = (x, y)^T$ 和相机坐标系下的空间坐标的关系求解所述手部深度图像中任一点的相机坐标系中的空间坐标 x_c, y_c 和 z_c ,图像坐标和相机坐标系下空间坐标的关系如下:

$$[0045] \quad z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix}, \text{其中 } z_c \text{ 为相机的光轴};$$

[0046] 在 z_c 已知的条件下,可以求出每一点对应 x_c 和 y_c ,利用相机坐标系下的空间坐标建立对应每一幅手部深度图像且包括关节点和掌心坐标的原始点云。

[0047] 为了在弥补手部深度图像特征纹理单一的缺点,在本实施例中,对训练样本的预处理重点增加了手势的三维空间特征。具体来说,三维空间特征优选包括根据手部深度图像生成的原始点云数据中包含的法向量和曲率。在获得对应每一幅手部深度图像且包括关节点和掌心坐标的原始点云后,首先计算根据手部深度图像形成的原始点云数据的法向量和曲率。原始点云数据的法向量可以通过拟合法计算,曲率可以通过抛物面拟合法计算。在拟合计算的过程中,为了提高运算的速率,使用kd-tree算法构建原始点云的树结构,并利用原始点云的树结构查找计算原始点云数据的法向量和曲率。

[0048] 在得到原始点云数据的法向量和曲率后,对手部深度图像、对应每一幅手部深度图像的法向量和曲率进一步进行去均值,利用方差归一化至 $[-1, 1]$,并进一步利用PCA算法将手部深度图像降维到 96×96 ,减少后续卷积神经网络的数据处理量,同时保证有效特征保留充分。

[0049] 获得归一化后的曲率、法向量和降维到 96×96 的手部深度图像后,基本完成了对训练图像的预处理。

[0050] 下一步利用预处理后的图像和数据训练CNN网络。在本实施例中,参见图2所示,搭建的CNN网络包括并行的第一卷积通道,第二卷积通道和第三卷积通道,其中第一卷积通

道、第二卷积通道和第三卷积通道的输入端分别输入预处理生成的对应手部深度图像的原始点云的法向量、曲率和手部深度图像,在并行的第一卷积通道、第二卷积通道和第三卷积通道中均包括三级卷积,如图2所示C1,C2,C3,在每一级卷积层后均跟随一层池化层,如图2所示P1,P2,P3,即采样层,第一卷积通道、第二卷积通道和第三卷积通道输出至全连接层。全连接层包括三级全连接层,如图2所示f1,f2,f3,其中优选第一级全连接层f1包括1024个神经元,第二级全连接层f2包括1024个神经元,第三级全连接层f3包括512个神经元。第三级全连接层f3输出包、对应手部深度图像中多个关节点,包括掌心在内的三维坐标3J,其中J代表关节数。如图2所示,在本实施例中,优选每一级卷积层的卷积核为 5×5 ,优选选择8个卷积核,池化层的池化核为 2×2 。在CNN网络中,激活函数为Relu函数。

[0051] 通过训练得到的CNN网络作为三维手势的特征提取器存储在VR头戴式设备的存储单元中。当VR头戴式设备运行时,通过VR头戴式及设备本身带有的深度摄像头采集含有三维手势的实时动作深度图像,对实时动作深度图像进行预处理,将实时动作深度图像中的手部深度图像和背景深度图像区分开,提取前部的深度图像,并对深度图像进行去噪。对深度图像进行降维,将图像分辨率降为 256×256 ,根据实时动作深度图像建立点云数据,利用点云数据计算法向量和曲率,并将实时动作深度图像、法向量和曲率输入至CNN网络的三个卷积通道的输入端,利用特征提取器识别实时动作深度图像中的三维手势,并生成识别出的三维手势中关节点和掌心的坐标,定位三维手势动作,并将识别结果输出至处理器,处理器对识别出的手势进行跟踪。

[0052] 本发明在对原始手势深度数据的预处理过程中对数据的法向量、曲率这两种三维空间特征信息进行了提取,并在横向尺度和纵向层深上做了改进。通过对采集数据三维空间描述信息的提取和利用,弥补了深度图像特征纹理单一的不足,强化了深度数据所具有的三维空间特征,三维手势无需做分类,因此本发明所提供的方法可以适用于对连续变动手势的追踪,同时,由于在采集数据中增加了三维空间的描述信息,因此可以最大可能的排除相机采集角度带来的误差。本发明提高了CNN训练模型的精确度,进而提高了VR头戴设备对手势跟踪的精确度。

[0053] 本发明同时公开了一种VR头戴式显示设备,采用如上述实施例所详细描述VR头戴式显示设备的手势跟踪方法。手势跟踪方法的具体数据采集、处理和识别过程参见上述实施例,在此不再赘述,本发明所公开的VR头戴式显示设备具有同样的技术效果。

[0054] 最后应说明的是:以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

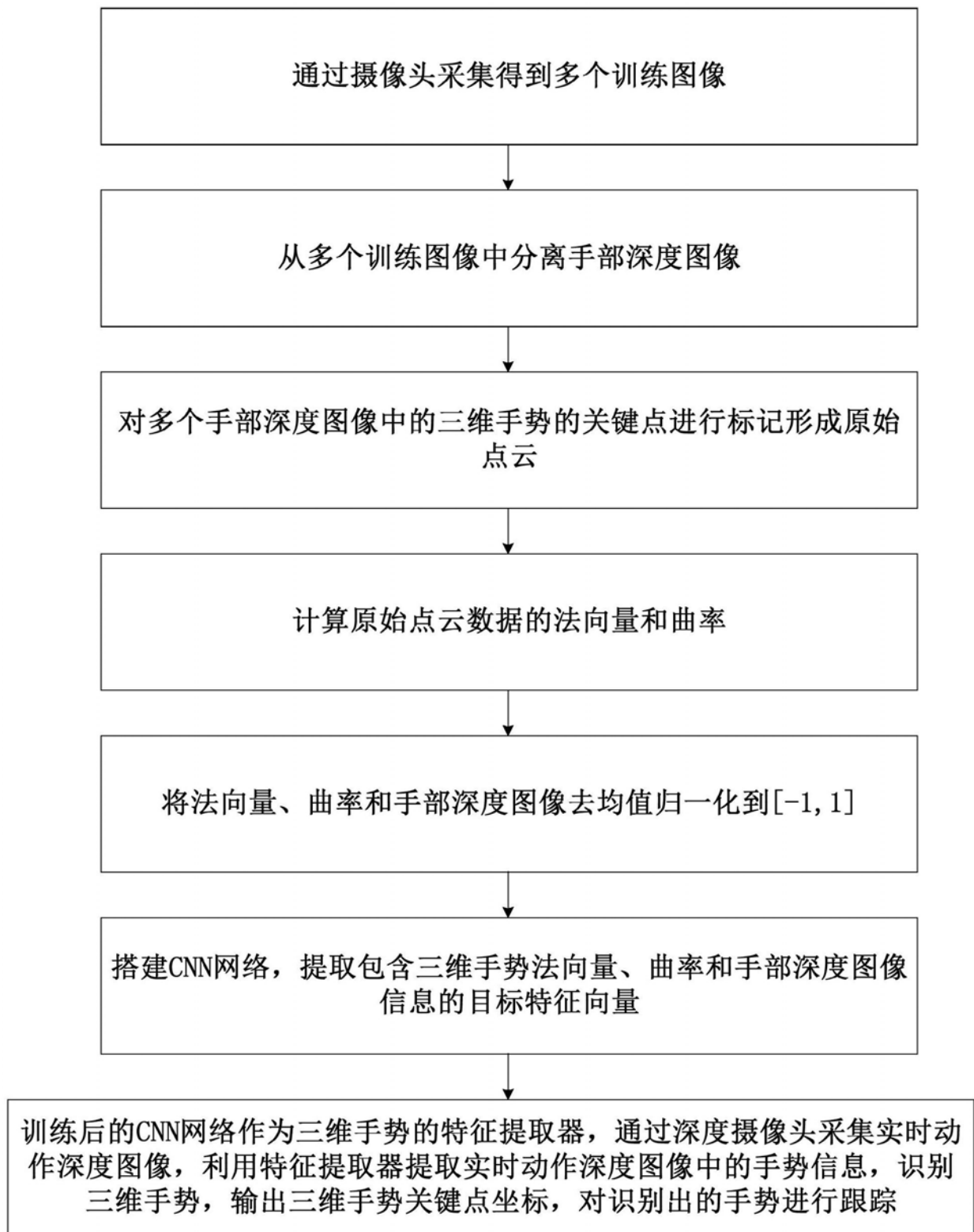


图1

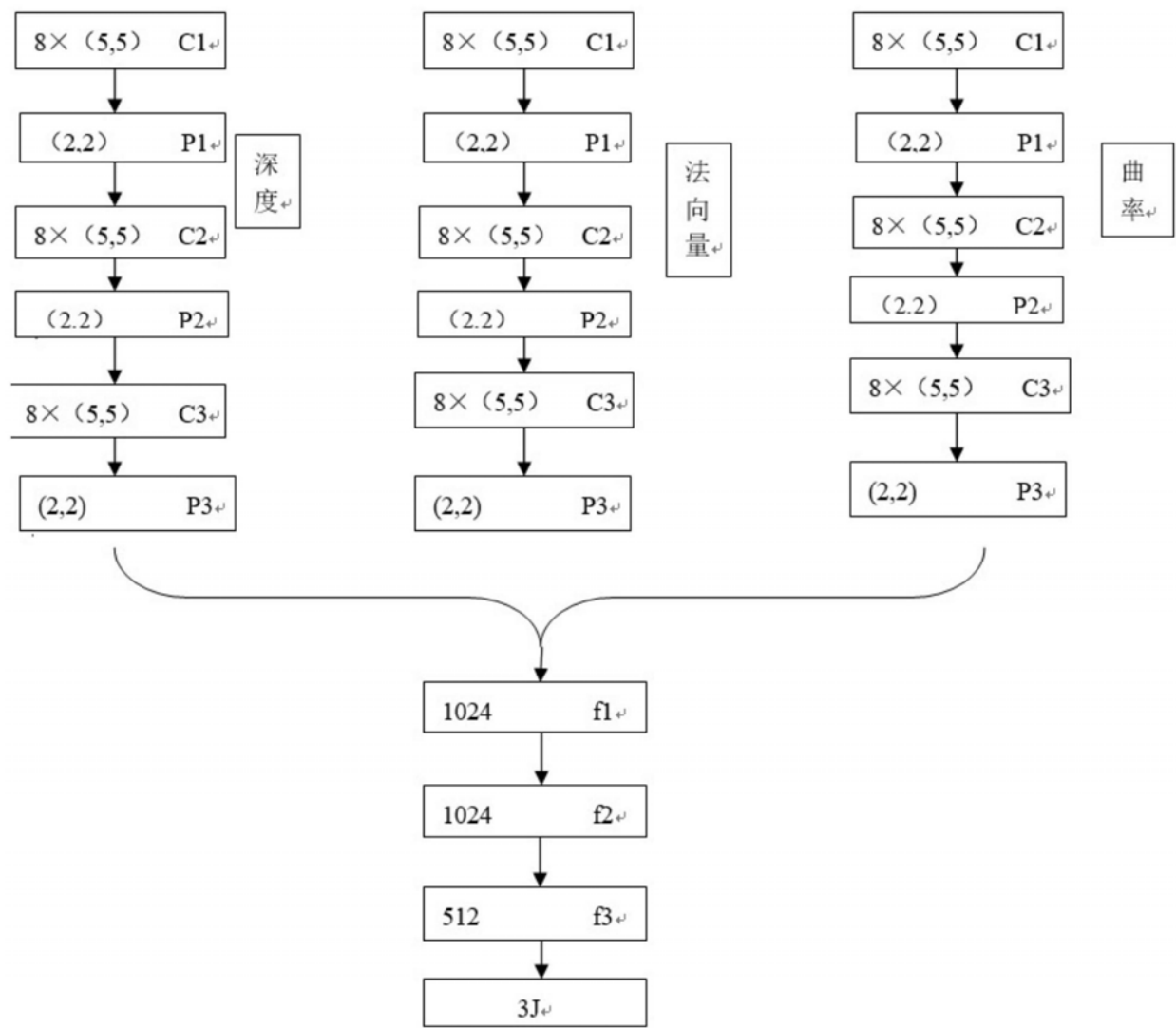


图2