



(12) 发明专利申请

(10) 申请公布号 CN 104317391 A

(43) 申请公布日 2015. 01. 28

(21) 申请号 201410494445. X

(22) 申请日 2014. 09. 24

(71) 申请人 华中科技大学

地址 430074 湖北省武汉市洪山区珞喻路
1037 号

申请人 富士通株式会社

(72) 发明人 葛俊锋 何源 叶林 刘伟 孙俊
黄文佳

(74) 专利代理机构 华中科技大学专利中心
42201

代理人 廖盈春

(51) Int. Cl.

G06F 3/01 (2006. 01)

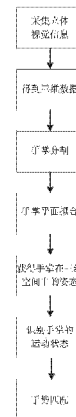
权利要求书2页 说明书5页 附图3页

(54) 发明名称

一种基于立体视觉的三维手掌姿态识别交互方法和系统

(57) 摘要

本发明公开了一种基于立体视觉的三维手掌姿态识别方法和系统,属于计算机视觉和人机交互技术领域。本发明包括:使用立体视觉信息采集设备进行数据的采集,得到包括深度信息在内的三维数据;通过得到的三维数据对手掌进行分割,然后运用平面拟合,获得手掌在三维空间中的姿态;根据跟踪识别的手掌姿态,匹配预定义的标准手势,进行具体的三维应用,实现人机交互。本发明通过识别手掌的姿态,使得识别的精度更高,更少的出现误识别的操作,同时解决了一般基于指尖识别的高运算量的问题,实现了更加实时的人机互动功能,保证了交互的准确性和易用性。



1. 一种基于立体视觉的三维手掌姿态识别交互方法,其特征在于,包括:

步骤 1 使用立体视觉信息采集设备采集图像作为立体视觉信息;

步骤 2 根据采集的所述立体视觉信息,获取其三维数据;

步骤 3 根据得到的所述三维数据,使用分割算法对手掌进行分割,分割得到所述手掌的点云以及手掌区域的轮廓信息,应用点云平面拟合方法,获得所述手掌在三维空间中的姿态;

步骤 4 根据所述姿态对所述手掌的运动状态进行识别,获得所述运动状态的特征值信息;

步骤 5 将得到的连续多帧的所述运动状态的特征值与预定义的标准手势的特征值时间序列进行匹配,得到与所述运动状态最接近的标准手势,将其作为手势匹配结果进行。

2. 如权利要求 1 所述的方法,其特征在于,所述步骤 1 中使用的所述立体视觉信息采集设备包括:深度摄像机、双目摄像机和多目摄像机。

3. 如权利要求 2 所述的方法,其特征在于,所述步骤 2 中所述双目摄像机获取所述三维数据具体包括以下子步骤:

(2-1) 使用 MATLAB 中的 Camera Calibration 工具箱分别对所述双目摄像机的左右摄像头单独进行内部参数标定,其中,定标参数包括所述左右摄像头的焦距、成像远点、畸变系数和双目相对位置;

(2-2) 根据得到的所述定标参数进行所述双目摄像机的外部参数标定;

(2-3) 使用所述双目摄像机采集左右图像,并利用 OpenCV 库和标定的所述外部参数对所述左右图像各自进行畸变校正,最后对两幅图像一起进行立体校正;

(2-4) 对立体校正后的两幅图像进行立体匹配,得到视差图,根据所述视差图计算得到所述三维数据。

4. 如权利要求 1-3 中任一项所述的方法,其特征在于,所述步骤 3 中对手掌进行分割具体包括以下步骤:

(a) 设待分割图像的原始灰度级为 M , 灰度值为 i 的像素点的个数为 n_i , 其中 $i = 0, 1, 2, \dots, M-1$, 对灰度值 i 出现的概率进行归一化: $P_i = \frac{n_i}{\sum n_i}$;

(b) 设分割阈值为 t , 将灰度值分为了两类, 每一类出现的概率分别为 $w_0 = \sum_{i=0}^t P_i$ 和 $w_1 = \sum_{i=t+1}^{M-1} P_i$, 每一类的平均灰度值分别为 $\mu_0 = \frac{\mu(t)}{w_0}$ 和 $\mu_1 = \frac{\mu_T(t) - \mu_0(t)}{1 - w_0}$, 其中

$$\mu(t) = \sum_{i=0}^t i \times P_i, \quad \mu_T(t) = \sum_{i=t+1}^{M-1} i \times P_i;$$

(c) 定义类间方差为: $\sigma^2 = w_0 \times (\mu_0(t) - \mu_1(t))^2 + w_1 \times (\mu_1(t) - \mu_1(t))^2$, 其中 t 从 0 到 $M-1$ 之间变化, 使得所述类间方差 σ^2 最大的值即为分割最佳阈值。

5. 如权利要求 1 所述的方法,其特征在于,所述步骤 3 还包括:

比较分割所获得的所述手掌区域与所述图像的比例是否小于预设的比例阈值,是则判断所述图像内不存在手掌。

6. 如权利要求1所述的方法,其特征在于,所述步骤3中使用的所述点云平面拟合算法具体包括以下步骤:

- (A) 以相同的概率初始化分割得到的每个手掌区域的三维点集;
- (B) 随机选取一组用于求解拟合平面参数的最小点集,计算相应的手掌拟合平面参数;
- (C) 计算所述三维点集中所有点到拟合平面的距离误差;
- (D) 根据每个点到所述拟合平面的误差大小调整其采样概率,并将所有点的采样概率归一化,每次更新点的采样概率时,参与本次平面拟合的点的采样概率保持不变;
- (E) 根据更新后的采样概率选取新的最小点集,依次迭代下去,直到每个点的采样概率不再变化。

7. 如权利要求1所述的方法,其特征在于,所述步骤4中对所述运动状态进行识别具体包括以下子步骤:

(4-1) 将手掌的所述姿态用三维向量 (x_1, y_1, z_1) 表示,相邻前一帧或后一帧中的手掌姿态用三维向量 (x_2, y_2, z_2) 表示;

(4-2) 手掌运动的速度向量 $V = (x_2, y_2, z_2) - (x_1, y_1, z_1)$;

(4-3) 通过是否存在手掌、手掌姿态信息、以及手掌的运动速度,可以判断手势的开始及结束,其中,设定一个速度阈值,当所述速度向量 V 的模大于所述速度阈值并维持若干帧时,判断是手势开始;当所述速度向量 V 的模小于所述速度阈值并维持若干帧时,判断是手势结束;如果检测到手掌,但所述手掌在若干帧内并没有运动,则判断是静态的手势。

8. 一种基于立体视觉的三维手掌姿态识别交互系统,其特征在于,包括:

三维数据采集单元,通过立体视觉信息采集设备采集图像作为立体视觉信息,获取其三维数据;

手掌三维姿态识别单元,通过得到的所述三维数据对手掌进行分割,分割得到所述手掌的点云后,应用点云平面拟合方法,获得所述手掌在三维空间中的姿态;

手势匹配单元,根据所述姿态对所述手掌的运动状态进行识别,获得所述运动状态的特征值信息,将得到的连续多帧的所述运动状态的特征值与预定义的标准手势的特征值时间序列进行匹配,得到所述运动状态最接近的标准手势,将其作为手势匹配结果;

人机交互单元,根据所述手势匹配结果,进行三维交互应用,实现人机交互。

9. 如权利要求8所述基于立体视觉的三维手掌姿态识别交互系统,其特征在于,所述预定义的标准手势包括:

单手以手腕为支点进行手掌的摆动,控制待交互物体的上下或左右移动;

单手在垂直于摄像机的平面内自由移动,控制所述待交互物体在平面内的移动;

单手在前后方向上水平移动,控制所述待交互物体的放大或缩小;

双手手掌相对进行移动,控制所述待交互物体的放大或缩小;

双手手掌相对进行旋转,控制所述待交互物体的旋转。

一种基于立体视觉的三维手掌姿态识别交互方法和系统

技术领域

[0001] 本发明属于计算机视觉和人机交互技术领域,更具体地,涉及一种基于立体视觉的三维手掌姿态识别方法和系统。

背景技术

[0002] 传统的基于鼠标、键盘、遥控器等人机交互技术虽然已经相对完善,但是在操作方式上并不人性化。且在某些场合,传统的鼠标键盘以及遥控器已经无法满足新的交互需求。

[0003] 随着科技的发展,基于视觉的人机交互技术已经成为了一个主流技术。这种技术一般是采用手势控制,而手势识别又可以分为二维和三维。二维感测技术可直接透过相机的 X 轴及 Y 轴影像进行手势判断,但由于缺乏深度信息,仅能判断简单的肢体动作,在某些交互场景显得不足。

[0004] 而三维手势交互方式,从信息表达能力上来看,能够表示复杂的多维关系。从交互方式上来看,三维人机交互方式模拟现实世界中人与物体的交互方式,更加清晰。但是现有的三维手势通常都还是基于对指尖检测识别手势进行匹配,无论从精度上还是适用性上都存在一定的局限性。

发明内容

[0005] 针对现有技术的以上缺陷或改进需求,本发明提供一种基于立体视觉的三维手掌姿态识别交互方法和系统,通过探测手掌在三维空间中的运动姿态,同时匹配预定义的手势,实现人机交互。

[0006] 为实现上述目的,按照本发明的一个方面,提供一种基于立体视觉的三维手掌姿态识别交互方法,包括以下步骤:

[0007] 步骤 1 使用立体视觉信息采集设备采集图像作为立体视觉信息;

[0008] 步骤 2 根据采集的所述立体视觉信息,获取其三维数据;

[0009] 步骤 3 根据得到的所述三维数据,使用分割算法对手掌进行分割,分割得到所述手掌的点云以及手掌区域的轮廓信息,应用点云平面拟合方法,获得所述手掌在三维空间中的姿态;

[0010] 步骤 4 根据所述姿态对所述手掌的运动状态进行识别,获得所述运动状态的特征值信息;

[0011] 步骤 5 将得到的连续多帧的所述运动状态的特征值与预定义的标准手势的特征值时间序列进行匹配,得到与所述运动状态最接近的标准手势,将其作为手势匹配结果进行。

[0012] 按照本发明的另一方面,提供一种基于立体视觉的三维手掌姿态识别交互系统,包括:

[0013] 三维数据采集单元,通过立体视觉信息采集设备采集图像作为立体视觉信息,获取其三维数据;

[0014] 手掌三维姿态识别单元,通过得到的所述三维数据对手掌进行分割,分割得到所述手掌的点云后,应用点云平面拟合方法,获得所述手掌在三维空间中的姿态;

[0015] 手势匹配单元,根据所述姿态对所述手掌的运动状态进行识别,获得所述运动状态的特征值信息,将得到的连续多帧的所述运动状态的特征值与预定义的标准手势的特征值时间序列进行匹配,得到所述运动状态最接近的标准手势,将其作为手势匹配结果;

[0016] 人机交互单元,根据所述手势匹配结果,进行三维交互应用,实现人机交互。

[0017] 总体而言,通过本发明所构思的以上技术方案与现有技术相比,具有以下有益效果:

[0018] 相较于现有基于指尖识别的技术,本发明通过识别手掌的姿态,使得识别的精度更高,更少的出现误识别的操作。同时解决了一般基于指尖识别的高运算量的问题,减少了运算的时间,实现了更加实时的人机互动功能。对比基于手套的手势识别方法,本发明只需要一个深度信息采集设备,成本更低,也更加自然,保证了交互的准确性和易用性。

附图说明

[0019] 图1为本发明基于立体视觉的三维手掌姿态识别交互方法的流程图;

[0020] 图2为本发明获得三维数据之前的结构框图;

[0021] 图3为本发明基于立体视觉的三维手掌姿态识别交互系统的结构框图;

[0022] 图4为本发明基于立体视觉的三维手掌姿态识别系统使用的示意图。

具体实施方式

[0023] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。此外,下面所描述的本发明各个实施方式中所涉及到的技术特征只要彼此之间未构成冲突就可以相互组合。

[0024] 图1所示为本发明基于立体视觉的三维手掌姿态识别交互方法的流程图,具体包括以下步骤:

[0025] 步骤1 采集图像作为立体视觉信息,立体视觉信息采集设备可包括:深度摄像机、双目摄像机、多目摄像机等。

[0026] 步骤2 根据采集的立体视觉信息,获取其三维数据。深度摄像机例如 TOF 摄像机、Kinect 等可以直接获取原始图像的三维数据,双目或者多目摄像机可以通过三维重建方法获取深度等三维数据。在本发明一个实施例中,采用双目摄像机作为深度采集设备。利用 Camera Calibration 工具箱,对双目摄像机进行内外部参数的标定,其中定标参数包括单目的焦距、成像远点、畸变系数和双目相对位置即旋转矩阵和平移向量。利用上述对双目摄像机标定的结果,可以使用 OpenCV 中的函数库分别对双目摄像机采集的左右视图进行消除畸变和行对准,使得左右视图的成像原点坐标一致、两摄像头光轴平行、左右成像平面共面、对极线行对齐。再使用极性约束,将图像的二维匹配问题降为一维搜索,通过立体匹配,得到视差图。其中立体匹配方法有很多,例如 BM 算法、SGBM 算法和 GC 算法。上述提到的算法速度依次递减,但是匹配得到视差的精度也依次提高。由于本发明最终是为了在视频中对手掌姿势进行匹配,因此优先采用速度快、效果能够达到要求的算法,即选择

BM 算法。具体 BM 算法参数的选择可通过实验获取。在获得了视差信息后,根据投影模型可以得到原始图像的深度信息和三维信息,即三维数据。在 OpenCV 中,直接调用其提供的 reprojectImageTo3D() 函数来完成三维信息的获取。在本发明另一实施例中,由于其他方式的深度信息采集设备例如深度摄像机可以直接得到三维数据,不需要进行上述处理。本发明步骤 2 的上述方法,均可以直接使用 OpenCV 中的函数实现。

[0027] 步骤 3 得到摄像机拍摄区域的立体视觉信息的三维数据后,采用聚类的思想,再综合使用阈值分割、区域分割等方法对手掌进行分割,分割得到手掌的点云以及手掌区域的轮廓信息,应用点云平面拟合方法,识别手掌在三维空间中的姿态。

[0028] 在本发明实施例中,可以综合使用多种分割算法进行手掌分割,例如自适应阈值分割、区域分割和边缘分割等。以选用 OSTU 自适应阈值分割算法进行分割为例,假设待分割图像的原始灰度级为 M ,其中灰度值为 i ($i = 0, 1, 2, \dots, M-1$) 的像素点个数为 n_i ,对灰度值 i 出现的概率进行归一化: $P_i = \frac{n_i}{\sum n_i}$; 设分割阈值为 t ,将灰度值分为了两类,每一

类出现的概率分别为 $w_0 = \sum_{i=0}^t P_i$ 和 $w_1 = \sum_{i=t+1}^{M-1} P_i$; 每一类的平均灰度值分别为 $\mu_0 = \frac{\mu(t)}{w_0}$ 和

$\mu_1 = \frac{\mu_T(t) - \mu_0(t)}{1 - w_0}$, 其中 $\mu(t) = \sum_{i=0}^t i \times P_i$, $\mu_T(t) = \sum_{i=t+1}^{M-1} i \times P_i$ 。则类间方差定义为: $\sigma^2 = w_0$

$\times (\mu_0(t) - \mu_1(t))^2 + w_1 \times (\mu_1(t) - \mu_1(t))^2$, 其中 t 从 0 到 $M-1$ 之间变化,使得该类间方差 σ^2 最大的值即为分割最佳阈值。这样就可以分割得到手掌所包含的点云以及手掌区域的轮廓信息。设定一个比例阈值 S_{min} , 例如 $1/10$, 如果分割所获得手掌区域与全图的比例小于比例阈值 S_{min} , 则判断不存在手掌。通过对手掌轮廓进行模板匹配分析,可以得到手指的位置、伸展状态,例如有几根手指伸展等。通过手指的朝向,可用于判断摆手等动作。分割得到手掌的点云后,应用点云平面拟合方法,获得手掌在三维空间中的姿态就是手掌的姿态。

[0029] 手掌平面拟合可以采用最小二乘拟合,但该方法容易受野值点影响,本发明提出一种概率引导的随机采样算法,可以有效克服野值点对手掌平面拟合的影响,具体包括以下步骤:

[0030] (A) 以相同的概率初始化经过分割得到的每个手掌区域的三维点集(即上述分割算法得到的点集);

[0031] (B) 随机选取一组用于求解拟合平面参数的最小点集,在本发明实施例中,最小点集通常为 3 个点,计算相应的手掌拟合平面参数;

[0032] (C) 计算该三维点集中所有点到该拟合平面的距离误差;

[0033] (D) 根据每个点到该拟合平面的误差大小调整其采样概率,并将所有点的采样概率归一化,即 $p(x_i) = p(x_i) / \sum p(x_i)$ 。假设原采样概率为 $p(x_i)$, 距离误差为 e_i , 则新的采样概率为 $p(x_i)' = p(x_i) + 1/e_i$, 为了防止含有野值点的平面将野值点的采样概率也同时提高,每次更新点的采样概率时,参与本次平面拟合的点的采样概率保持不变;

[0034] (E) 根据更新后的采样概率选取新的最小点集,依次迭代下去,直到每个点的采样概率不再变化。

[0035] 步骤 4 根据上述得到的手掌在三维空间中的姿态对手掌运动状态进行识别,获得

手掌运动状态的特征信息。在本发明实施例中,手掌运动状态可以根据相邻两帧或者多帧的手掌姿态的变化来获取,以相邻两帧计算方法为例,具体包括以下子步骤:

[0036] (4-1) 将手掌的姿态用三维向量 (x_1, y_1, z_1) 表示,相邻前一帧或后一帧中的手掌姿态用三维向量 (x_2, y_2, z_2) 表示;

[0037] (4-2) 手掌运动的速度向量 $V = (x_2, y_2, z_2) - (x_1, y_1, z_1)$;

[0038] (4-3) 通过是否存在手掌、手掌姿态信息、以及手掌的运动速度,可以判断手势的开始及结束。其中,在本发明实施例中,设定一个速度阈值 V_{min} ,当速度向量 V 的模大于 V_{min} 并维持若干帧时,判断是手势开始;当速度向量 V 的模小于速度 V_{min} 并维持若干帧时,判断是手势结束;如果检测到手掌,但手掌在若干帧内并没有运动,则判断是静态的手势。

[0039] 手掌运动状态的特征值由每帧手掌姿态向量、速度向量等构成,对手掌运动状态的特征值进行归一化处理,消除尺度因素对特征的影响。

[0040] 步骤 5 将得到的连续多帧的手掌运动状态的特征值传入已训练好的隐马尔科夫模型中,与预定义的标准手势的特征值时间序列进行匹配,得到待识别手势最接近的标准手势,将其作为手势匹配结果。

[0041] 在本发明实施例中,预定义三维交互手掌姿势包括以下例子,但不以此为限:单手以手腕为支点进行手掌的摆动,控制待交互物体的上下或左右移动;单手在垂直于摄像头的平面内自由移动,控制待交互物体在平面内的移动;单手在前后方向上水平移动,控制待交互物体的放大或缩小;双手手掌进行相对移动,控制待交互物体的放大或缩小;双手手掌相对进行旋转,控制待交互物体的旋转。根据上述预定义的三维交互手掌姿态,建立一套基于手掌姿势跟踪的交互模型。

[0042] 图 2 所示为本发明获得三维数据之前的数据处理流程图。在本发明实施例中,以双目摄像机为例,具体包括以下子步骤:

[0043] (2-1) 使用 MATLAB 中的 Camera Calibration 工具箱分别对左右摄像头单独进行内部参数标定,其中定标参数包括左右摄像头的焦距、成像远点、畸变系数和双目相对位置即旋转矩阵和平移向量;

[0044] (2-2) 根据得到的左右摄像头标定参数进行双目摄像机外部参数标定;

[0045] (2-3) 使用双目摄像机采集左右图像,并利用 OpenCV 库和标定的外部参数对左右图像各自进行畸变校正,最后对两幅图像一起进行立体校正;

[0046] (2-4) 对立体校正后的两幅图像进行立体匹配,得到视差图,即可根据视差图计算得到三维数据。

[0047] 图 3 所示为本发明基于立体视觉的三维手掌姿态识别交互系统的结构框图,具体包括以下单元:

[0048] 三维数据采集单元,通过立体视觉信息采集设备采集图像作为立体视觉信息,获取其三维数据;

[0049] 手掌三维姿态识别单元,通过得到的三维数据对手掌进行分割,分割得到手掌的点云后,应用点云平面拟合方法,获得手掌在三维空间中的姿态;

[0050] 手势匹配单元,根据姿态对手掌的运动状态进行识别,获得运动状态的特征值信息,将得到的连续多帧的运动状态的特征值传入已训练好的隐马尔科夫模型中,与预定义

的标准手势的特征值时间序列进行匹配,得到待识别手势最接近的标准手势,将其作为手势匹配结果;

[0051] 人机交互单元,根据手势匹配的结果,进行三维交互应用,实现人机交互。

[0052] 图4所示为本发明基于立体视觉的三维手掌姿态识别系统使用的示意图,展示了如何使用本系统。显示器里面是待控制的物体,手掌在摄像头平面内进行移动,实现人机交互。具体的,单手以手腕为支点进行手掌的摆动,控制待交互物体的上下或左右移动;单手在垂直于摄像头的平面内自由移动,控制待交互物体在平面内的移动;单手在前后方向上水平移动,控制待交互物体的放大或缩小;双手手掌进行相对移动,控制待交互物体的放大或缩小;双手手掌相对进行旋转,控制待交互物体的旋转。

[0053] 本领域的技术人员容易理解,以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,均应包含在本发明的保护范围之内。

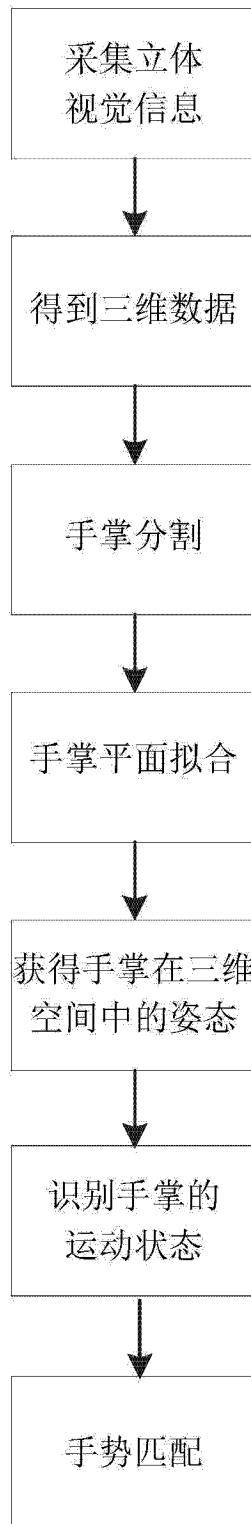


图 1

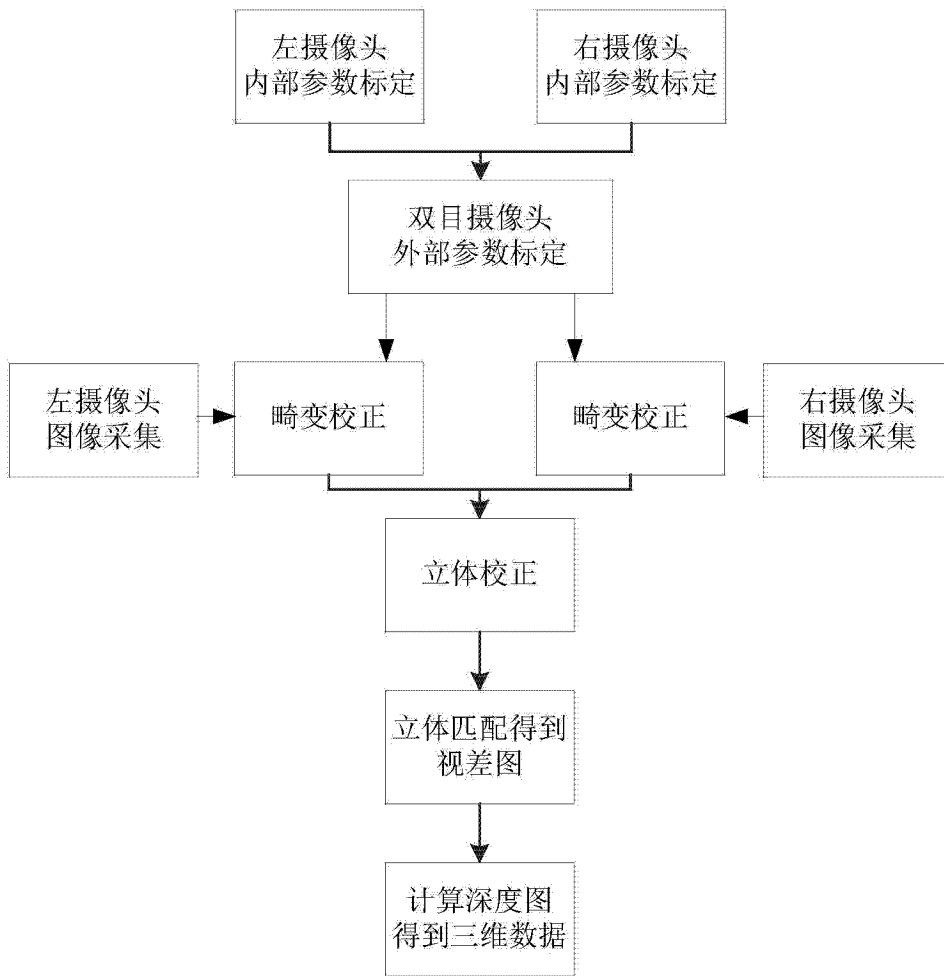


图 2



图 3

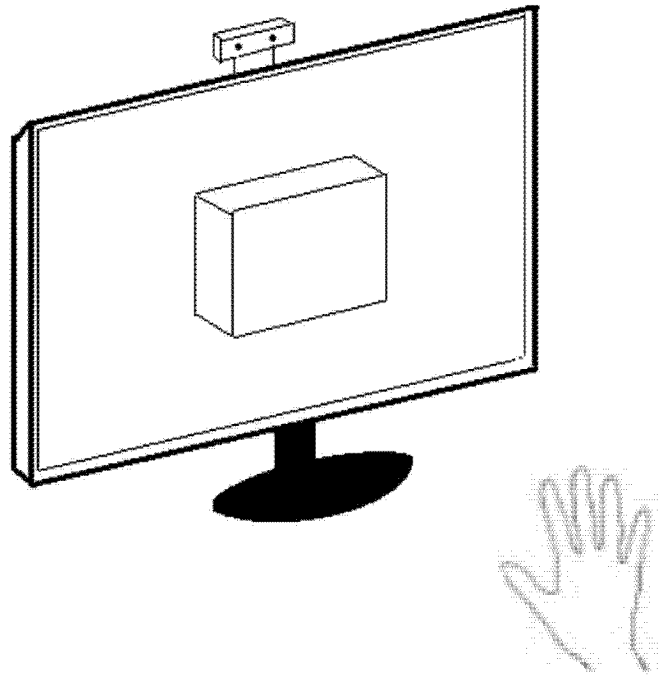


图 4