

基于 RGB-D 图像的头部姿态检测

刘振宇 关 彤

(沈阳工业大学信息科学与工程学院 沈阳 110870)

摘 要 在经颅磁刺激治疗过程中,准确且快速地检测人体头部姿态至关重要。针对基于二维彩色图像的头部姿态估计对环境、姿态敏感的问题,文中提出了一种联合彩色图像与深度图像的头部姿态检测方法。通过彩色图像检测人脸特征点的二维位置信息,结合深度信息定义三维头部坐标系;然后在现有的 ICP 点云配准算法的基础上,提出了一种粗配准方法。通过计算待检测头部点云与标准头部点云之间坐标系的变换关系得到初始位姿参数,以防止点云配准陷入局部最优局面。实验表明,该算法能够在光源均匀且充足的诊疗室环境中准确地检测人体头部姿态,提高头部姿态角度大时姿态估计的鲁棒性。

关键词 头部姿态检测,RGB-D 图像,人脸特征点,三维头部坐标系,点云配准

中图分类号 TP391.41 **文献标识码** A

Head Posture Detection Based on RGB-D Image

LIU Zhen-yu GUAN Tong

(School of Information Science and Engineering,Shenyang University of Technology,Shenyang 110870,China)

Abstract In the process of transcranial magnetic stimulation treatment,it is important to accurately and quickly detect the posture of the human head. Aiming at the problem that the head pose estimation based on two-dimensional color image is sensitive to the environment and posture,a head posture detection method combining both color image and depth image was proposed. The two-dimensional position information of the face feature points is detected by the color image, and the three-dimensional head coordinate system is defined by combining the depth information;Then,based on the existing ICP point cloud registration algorithm,a coarse registration method was proposed. The initial pose parameters are obtained by calculating the transformation relationship between the coordinate system of the head cloud,to be detected and the standard head point cloud,to protect the point cloud registration from falling into local optimum. Experiments show that the algorithm can accurately detect the head posture of the human body in a consulting room environment where the light source is uniform and sufficient,and improve the robustness of the attitude estimation when the head posture angle is large.

Keywords Head posture detection,RGB-D image,Face feature points,Three-dimensional head coordinate system,Point cloud registration

1 引言

经颅磁刺激仪通过磁信号无衰减地透过颅骨刺激大脑相应的神经区域,能够有效地治疗神经心理疾病、儿童脑瘫等。通过检测当前人体头部相对于摄像机的位姿,结合三维头部坐标系与医学中颅骨坐标系的转换关系,就可得到头部各穴位及神经区域的位姿,指挥装有磁刺激线圈的机械手臂跟随患者头部的运动自动定位到需要治疗的头部区域,实现治疗过程的智能化。本文提出的头部姿态检测技术将应用于经颅磁刺激仪的人机交互系统。

在人体头部中,人脸的特征信息最为丰富且易于采集及检测,因此头部姿态检测大多基于人脸图像信息。Murphy^[1]于 2009 年发表了头部姿势估计研究领域的第一篇综述性论文,并从方法学的角度将头部姿态检测方法分为 8 类,包括模

板匹配法、检测器阵列法、柔性模板法^[2-3]、跟踪法^[4-5]、回归法^[6-7]、几何计算法^[8]等。该论文综述的文献大多基于 RGB 图像检测头姿,缺乏对基于深度图像和三维数据的头部姿势检测方法的总结和比较。根据获取人脸图像方式的不同,也可将头部姿态检测方法分类为基于二维彩色图像、基于深度信息^[9]以及基于 RGB-D 图像^[10-12] 3 种^[13]。基于二维彩色图像的头部姿态检测法对硬件没有特殊要求,设备成本低,是传统的主流方法。例如,闵秋莎等^[14]利用 Hough 圆检测方法定位眼睛和鼻孔,将定位结果与其正脸姿态时的相对位置信息进行对比,从而得出不同头部姿态的粗估计。该文头姿检测方法的精度取决于图像预处理过程中人脸特征定位的精度,定位算法需要训练大量的图片。实验证明,该头部姿态检测方法的平均正确率可达 93.53%,在头部左右转和仰视角时的误差大于平均误差。因为头部在向左右转或仰视角度过

本文受辽宁省 2018 年自然科学基金(20180520022)资助。

刘振宇(1973—),男,博士,教授,主要研究方向为视觉伺服、模式识别;关彤(1994—),女,硕士生,主要研究方向为视觉伺服、图像处理, E-mail:18842414590@163.com(通信作者)。

大时,特征点易缺失,使头部姿态估计发生误判。相比 RGB 图像,深度图像包含头部的三维信息,能有效克服在逆向问题求解中的病态性,减少光照、阴影、遮挡等因素对头部姿态检测的影响,具有更高的鲁棒性和准确性^[15],且实时性好。Derkach 等人基于深度图像设计了一个结合了基于特征(包括几何估计和外观估计)和基于字典两种不同算法的头部姿态估计系统,这两种估计方法相结合产生的结果优于每个单独的方法,并且其准确率高于基于字典的估计方法。但是在实验中,大约 9% 的情况下扫描检测不到特征,只有基于字典的头部姿态估计是可用的。该系统在 SASE^[16] 数据库上进行头部姿态估计的准确度为 91.1%^[17]。李成龙等人采用 Kinect 相机获取深度图像,在使用卡尔曼滤波预测头部位置的同时,改善了随机回归森林算法检测的速度与准确率,最后结合预测值和测量值得到准确的头部姿态^[18]。与基于二维彩色图像的头姿检测方法相比,该方法可以很好地处理光照变化、遮挡和自遮挡等问题。Meyer 等^[2] 利用商品深度相机精确估计三维头部姿态,文中通过配准测量的深度数据与可变形人脸模型来估计姿态,采用粒子群优化算法(PSO)和迭代最近点算法(ICP)相结合的方法,使包含三维配准和二维重叠项的代价函数最小化。目前,通过消费级深度相机(例如微软 Kinect、英特尔实感深度摄像头系列等)获得的深度图像分辨率低,且有较为严重的噪声和信息缺失。RGB 数据在检测阶段可以为深度数据补充颜色及纹理信息,但它可能会降低 3D 信息所固有的对光照的鲁棒性。因此,我们认为基于 RGB-D 图像的头部姿态检测方法具有一定的优越性,研究的核心问题是如何更有效地结合彩色信息与深度信息^[18]。

本文提出了一种基于 RGB-D 图像检测头部姿态的方法,考虑实际应用中的需要,将头部姿态检测系统分为注册零姿态三维头部点云模型和头部姿态实时检测两部分。注册阶段在可控条件下进行,在高分辨率的彩色图像上检测人脸特征点,结合特征点的深度信息构建精度高的三维头部坐标系;实时检测阶段在非可控条件下进行,将头部坐标系粗配准计算与 ICP 点云配准算法相结合,得到待检测头部的精准头部姿态参数。

2 RGB-D 图像预处理

英特尔实感 SR300 摄像头(Intel® RealSense™ Camera SR300,SR300)是前置近距离集成 3D 空间深度和 2D 摄像头模块。3D 深度模块通过结构光原理和三维测距获取物体的深度信息,将深度数据可视化深度图像;2D 摄像头模块通过传统彩色摄像头获得含有丰富色彩信息和纹理信息的 RGB 图像。深度图像易受物体表面材质或遮挡等因素的影响,造成深度信息丢失和空洞;在观察深度相机 30 帧/s 的连续成像时,容易发现深度图像的一些细节部分在跳动,这些跳动的变化是由深度图像空洞和随机噪声共同影响产生的。综上所述,在人脸特征点检测、生成点云和点云配准之前,应对 RGB-D 数据进行针对性的预处理,以改善数据质量。

2.1 RGB-D 图像对齐

RGB-D 数据是将同一时刻下采集的深度数据和彩色数据对齐和融合的结果。传统摄像头和深度摄像头的安装位置不同,获得的图像大小也不同。本文后续算法要求 RGB-D 4 个通道对齐,即每一个三维点的颜色信息与深度信息一一对应。

假设深度图像某一像素点 (x,y) 的投影坐标为 p_{ir} ,其 z 值表示深度信息,单位为 mm。物体表面上一点在深度相机坐标系下的空间坐标为 P_{ir} ,深度相机内参为 H_{ir} ,则变换关系如式(1)所示:

$$\begin{cases} p_{ir} = H_{ir} \cdot P_{ir} \\ P_{ir} = H_{ir}^{-1} \cdot p_{ir} \end{cases} \quad (1)$$

同样,假设 P_{rgb} 为物体上同一点在彩色相机坐标系的空间坐标,该点在彩色图像上的投影坐标为 p_{rgb} 。 P_{rgb} 与 P_{ir} 之间的关系可通过式(2)表示:

$$P_{rgb} = R P_{ir} + T \quad (2)$$

彩色相机坐标系下某一像素点的三维坐标通过内参矩阵 H_{rgb} 变换,可得到其对应的 RGB 图像像素坐标:

$$p_{rgb} = H_{rgb} P_{rgb} \quad (3)$$

P_{ir} 和 P_{rgb} 均为齐次坐标。假设有点 P 空间坐标,那么:

$$\begin{cases} P_{ir} = R_{ir} P + T_{ir} \\ P_{rgb} = R_{rgb} P + T_{rgb} \end{cases} \quad (4)$$

在式(4)的第一式中,将 P 用 R_{ir} , P_{ir} , T_{ir} 表示,代入式(2)可得:

$$P_{rgb} = R_{rgb} R_{ir}^{-1} P_{ir} + T_{rgb} - R_{rgb} R_{ir}^{-1} T_{ir} \quad (5)$$

将上式进行整理,最后可得:

$$\begin{cases} R' = R_{rgb} R_{ir}^{-1} \\ T' = T_{rgb} - R_{rgb} R_{ir}^{-1} P_{ir} \end{cases} \quad (6)$$

在同一场景下,通过标定获取深度相机和彩色相机的内外参数矩阵,再由上式计算即可将 RGB 图像与 D 图像对齐。

2.2 深度图预处理

SR300 通过红外测距获得物体的深度信息,在物体的边缘处,反射信息往往发生剧烈的变化,使深度图像中物体的边缘模糊;空洞区域的反射光线不能被接收器接收,使深度信息不连续或产生图像空洞^[19]。深度数据通过相机标定原理转换为点云数据,深度图像预处理的效果直接影响点云的质量。如果深度图像中物体边缘受背景影响变得模糊,当转化为点空间坐标后,物体边缘就会出现变形。因此与传统的图像滤波要求相比,对深度图像进行滤波时,更加需要确保图像的边缘信息和细节信息,即对保边去噪的要求较高。

Tomas 和 Manduchi 提出的双边滤波算法是一种用于保边去噪的经典图像滤波算法,该算法结合高斯滤波算法和基于灰度相似度的领域均值滤波算法的优点,兼顾了滤波领域中各像素的空间距离和像素差两个重要影响因素^[20]。双边滤波的权值函数 w 由空间域函数 d 和值域函数 r 两部分构成,表达式如下:

$$\begin{cases} w(m,n) = d \times r \\ d(m,n) = e^{-\frac{(m-x)^2 + (n-y)^2}{2\sigma_d^2}} \\ r(m,n) = e^{-\frac{\| (m-x)^2 + (n-y)^2 \|}{2\sigma_r^2}} \end{cases} \quad (7)$$

其中, x,y 为目标像素坐标,即双边滤波模板中心坐标, m,n 为整个滤波领域中的像素坐标点。空间域函数 $d(m,n)$ 可以理解为标准差为 σ_d 的高斯滤波器,与滤波模板中心目标像素的欧氏距离越小,其权值越大;值域函数 $r(m,n)$ 是一个单调递减函数,像素差值越大, r 值越小。双边滤波的权值函数形式如式(8)所示:

$$w(m,n) = \exp\left(-\frac{(m-x)^2 + (n-y)^2}{2\sigma_d^2} - \frac{\| f(m,n) - f(x,y) \|^2}{2\sigma_r^2}\right) \quad (8)$$

由式(8)可知,在进行双边滤波时,不仅要确定滤波模板的大小,还要定义空间域标准差和值域标准差两个参数。空间域标准差越大,图像在像素变化较为平缓的区域变得越模糊;值域标准差的值越小,保边效果越好。

深度图像由深度数据可视化而来,通过灰度值的大小表示物体与相机之间的距离关系。通过设定合适的阈值 T ,与图像中每一个像素的灰度值相比较,灰度值大于 T 时认定其为背景,将灰度值设为 0;灰度值小于 T 则认定其为感兴趣区域,保留灰度值^[21]。

$$g(x,y)=\begin{cases} 0, & f(x,y)>T \\ f(x,y), & f(x,y)\leq T \end{cases} \quad (9)$$

2.3 点云计算

点云数据根据相机标定原理^[22],由深度数据转换而来,深度相机的成像原理如图 1 所示。我们定义深度相机的视点坐标系为头部姿态检测系统坐标系。

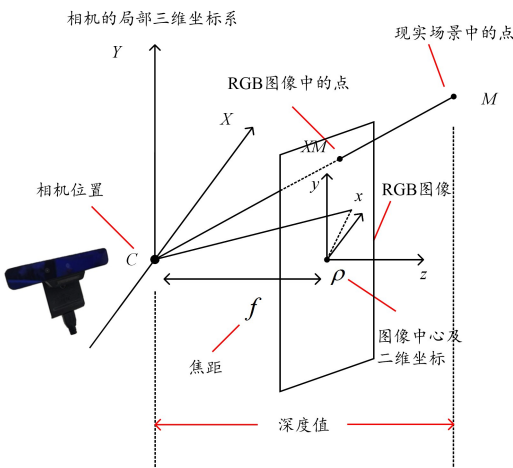


图 1 深度相机成像原理示意图

假设深度相机内参数矩阵为 $\mathbf{K}$,深度图像中的像素点 (i, j) 的深度值 $dep(i, j)$ 变换到深度相机坐标系下的三维点云坐标由式(10)实现:

$$\begin{bmatrix} x_{(i,j)} \\ y_{(i,j)} \\ z_{(i,j)} \end{bmatrix} = \left(dep(i,j) \cdot \left(\mathbf{K}^{-1} \cdot \begin{bmatrix} i \\ j \\ 1 \end{bmatrix} \right)^T \right)^T \quad (10)$$

其中, $\mathbf{K}^{-1}$ 是深度相机内参数矩阵的逆矩阵。XYZ 型点云数据直接由深度数据就可以得到,XYZRGB 型点云数据需要通过 RGB-D 数据进行转换。图 2 为在摄像机坐标系下的人脸点云图像。

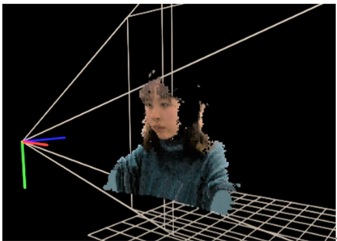


图 2 人脸点云图像

3 三维人脸特征点检测

Çelikütan 等人在关于人脸特征点检测的综述文章^[23]中概述了人脸特征点检测算法在过去十年中的进展以及未来的发展趋势。在面部处理文献中,我们将人脸特征点定义为人

脸面部的关键区域位置。例如,眼角、嘴角、鼻尖、眉毛这些使用初级的图像特征(方向梯度直方图、角度或尺度不变特征变换(SIFT)^[24])就可以直接检测得到的标志被称为主要标志(基准标志);鼻孔、下巴、脸颊轮廓、唇或眉毛中点、眼睑等的检测以基准标志为指导,被称为二级标志。人脸特征点被广泛地应用到面部表情分析、三维人脸重建、配准、人脸跟踪、头部姿态检测等领域。

三维特征点通过注册彩色图像和深度图像获得,本文把头部姿态粗检测转换为基于人脸特征点的三维头部坐标系变换问题,即寻找实时检测的人脸点云和零姿态三维人脸模型点云的姿态变换。特征点信息提取的准确性直接影响了姿态检测的可靠性。SR300 捕捉的彩色图像像素最高达 1080 像素,故为了得到点云间准确的特征匹配,首先从彩色图像中提取人脸特征点,获得其 x, y 空间的特征信息;再通过对齐的深度图像得到人脸特征点的深度信息。

首先,在 RGB 图像上基于 HOG 及线性分类器建立人脸检测器。然后,在检测到的人脸区域内检测特征点。在训练样本集中对人脸图像特征点进行标注,然后使用 Kazemi 等^[25]提出的 ERT(ensemble of regression trees)级联回归算法,实现人脸眼部、鼻子、嘴部等 68 个特征点位置的定位。

为避免光源直射造成人脸局部区域过亮或有阴影,本文将经颅磁刺激仪的实验环境设置为光源充足且均匀的室内,以确保人脸特征点检测的精度。针对单轴方向的姿态变化,在头部正常旋转角度范围内进行人脸特征点检测实验,部分姿态下实现的人脸 68 个特征点的检测效果如图 3 所示。

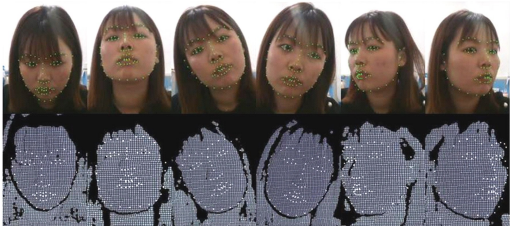


图 3 不同姿态图像中的人脸特征点检测

4 三维头部坐标系

物体的方位可由固接于该物体的坐标系来描述。要完全描述刚体在空间的位姿(位置与姿态),通常将物体 B 与某一坐标系 $\{B\}$ 相固接。 $\{B\}$ 的坐标原点一般选在物体 B 的特征点上,如质心等。相对于参考系 $\{A\}$,坐标系 $\{B\}$ 的原点位置和坐标轴的方位可由式(11)表示^[26]:

$$\{B\} = \{ {}^A_B R \quad {}^A P_{B_O} \} \quad (11)$$

本文假设人体头部是一个刚体,基于面部的特征点建立三维头部坐标系,用于头部姿态初始姿态的求解。

4.1 定义三维头部坐标系

建立三维标准头部坐标系的过程如下:

首先,检测人脸特征点,将其反投影到点云空间得到特征点的三维坐标。

然后,在三维人脸上选取 3 个稳定的面部标志来建立三维头部坐标系。人脸的内外眼角以及鼻子下端的中间点是面部的基准标志。为提高检测的准确性,分别取左右眼内外眼角的中点(记为点 M_1, M_2)和鼻子下端点(记为点 N)作为建立三维头部坐标系的基准点。

最后,定义原点及坐标轴。过点 N 做直线 M_1M_2 的垂线交其于点 O ,并令点 O 为头部坐标系原点;连接 $\vec{NO}$,规定下巴到眉心方向为 Z 轴正方向;连接 $\vec{OM_2}$,规定人脸从右眼到左眼方向为 Y 轴正方向;根据右手坐标系定则,确定头部点云模型的上方为 X 轴正方向。归一化 x, y, z 轴,计算过程如式(12)所示:

$$\begin{cases} \vec{o_y} = \frac{\vec{OM_2}}{\|\vec{OM_2}\|_2} \\ \vec{o_z} = -\frac{\vec{ON}}{\|\vec{ON}\|_2} \\ \vec{o_x} = \vec{o_y} \times \vec{o_z} \end{cases} \quad (12)$$

在人脸点云上建立的三维头部坐标系与 SR300 摄像机坐标系如图 4 所示。

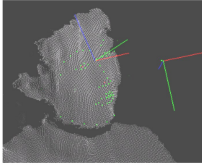


图 4 三维头部坐标系示意图

4.2 三维标准头部点云模型注册及点云优化

注册标准三维头部点云模型并规定其姿态为零,即绕 x, y, z 三轴的欧拉角均为 0,将其作为检测其他头部姿态的基准。为了提高配准精度,在使用标准点云模型之前对其点云数据进行优化处理^[27]。

注册阶段,使人脸正面对摄像头,采集 RGB-D 图像并生成零姿态头部点云模型。在建立三维头部标准点云模型的过程中,填充空洞是点云优化的首要任务。三维点云的本质仍是深度信息,可以通过直接处理深度信息来达到点云优化的效果^[19]。

多帧均值滤波算法能在去噪的同时对孔洞进行修补^[27]。在同样的条件下对同一场景拍摄 m 张图像,在理想情况下每张图像 $g(x, y)$ 都是完全一致的(像素的个数、排列、像素值的大小都完全相同),但受到加性噪声 $n_i(x, y)$ 的影响, m 张含噪图片的数学表示为:

$$f_i(x, y) = g(x, y) + n_i(x, y), i = 1, 2, \dots, m \quad (13)$$

多帧均值滤波的思想是基于每张图像的有效信号完全一致,噪声相互独立,且噪声 $n_i(x, y)$ 的均值都为 0 的假设之上的,利用 m 张图像的算术平均值代替理想图像 $g(x, y)$,具体计算如式(14)所示:

$$\begin{aligned} \bar{f}(x, y) &= \frac{1}{m} \sum_{i=1}^m f_i(x, y) \\ &= g(x, y) + \frac{1}{m} \sum_{i=1}^m n_i(x, y) \\ &\approx g(x, y) \end{aligned} \quad (14)$$

对 m 幅信号相同而噪声不同的图像求算术平均,因为噪声的均值为 0,所以含噪图像算术平均的数学期望就等于信号,同时算术平均后的方差减小为 $1/m$ 。因此,该算法在抑制噪声的同时,对输入图像没有进行任何平滑处理。

5 姿态检测

实时检测过程的关键部分,就是将待检测头部点云与三

维标准头部模型点云进行配准。若两点云有一定的重叠度(一般特征比较明显的点云重叠部分超过 60% 时),即可得到正确的配准结果。在本文中,自定义的零姿态三维头部标准模型与待检测三维人脸点云都有不同程度的重叠。若直接使用精配准算法,则容易陷入局部最优解,而得不到全局最优。因此,需要利用粗配准算法,通过点云的一些特征信息估计出初始配准参数,为精配准提供精确配准一个迭代的起始位置,从而增大收敛到最佳结果的概率。

针对上述点云数据配准中存在的问题,本文将二维图像上检测的人脸特征点映射到三维点云上,为点云增加了稳定的特征,并基于此面部特征建立三维头部坐标系,通过计算待检测头部坐标系与标准三维头部坐标系的相对位置关系得到初始配准参数,将两片点云在全局范围内的相对位置变换到 ICP 算法的收敛域内,具体流程如图 5 所示。

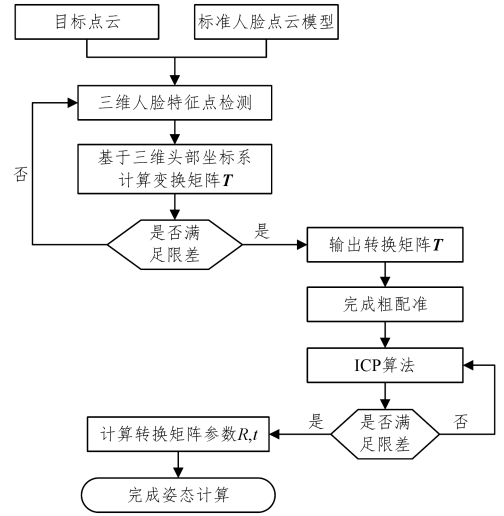


图 5 点云配准流程

5.1 粗配准

设零姿态头部模型点集为 P , 姿态待配准头部点集为 Q 。图 6 为点云粗配准示意图。

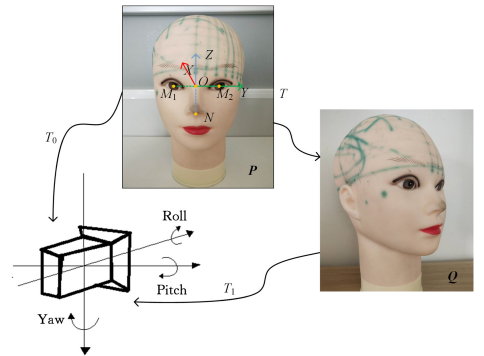


图 6 粗配准计算示意图

设头部 P 和 Q 中的各点在空间中齐次坐标变换矩阵为 T , 将其分解为旋转和平移两种运动:

$$T = \begin{bmatrix} {}^P_R & {}^P_t Q_0 \\ 0 & 1 \end{bmatrix} \quad (15)$$

在图 6 中,用旋转矩阵 P_R 描述头部坐标系 $\{Q\}$ 相对于 $\{P\}$ 的姿态。同一点 m 在两个三维头部坐标系 $\{P\}, \{Q\}$ 中的描述 ${}^P m$ 和 ${}^Q m$ 具有如下变换关系:

$${}^P m = {}^P_R {}^Q m \quad (16)$$

上式为坐标旋转方程。对旋转的表示可分解绕 x, y, z 3 个坐标轴的欧拉角, 分别表示为 α, β 和 γ 。头部各点以 ZYX 的顺序依次左乘旋转矩阵, 则其矩阵表示为: ${}^0R = R_x \cdot R_y \cdot R_z$ 。其中:

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{bmatrix} \quad (17-1)$$

$$R_y = \begin{bmatrix} \cos \beta & 0 & \sin \beta \\ 0 & 1 & 0 \\ -\sin \beta & 0 & \cos \beta \end{bmatrix} \quad (17-2)$$

$$R_z = \begin{bmatrix} \cos \gamma & -\sin \gamma & 0 \\ \sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (17-3)$$

设零姿态三维头部坐标系 $O-XYZ$ 与摄像机坐标系 $O_C-X_C Y_C Z_C$ 的空间齐次变换矩阵为 T_0 , 由于 $O_C = (0, 0, 0)$, 因此:

$$t_0 = O_C - O = O_C \quad (18)$$

$$R_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} (ox, oy, oz)^{-1} \quad (19)$$

待配准点云 Q 相对于摄像机坐标系的位置关系 T_1 的计算过程与 T_0 相同。由图 6 可知: $T = T_1^{-1} \cdot T_0$ 。 T 为粗配准结果, 使得 $TP = Q$ 。

5.2 精配准

ICP 算法由计算机视觉研究者 Besl 和 McKay^[28] 于 1992 年提出, 是目前最经典且较为成熟的点云配准方法。ICP 算法是一个基于最小二乘法的迭代过程, 可以使两个待配准的点云在距离最小和的目标函数下迭代收敛到局部最小值, 在整体上达到某种度量准则下的最优配准。

假设零姿态头部点云为源数据集 S , 待检测姿态点云为目标数据集 G , 基于粗配准结果更新目标数据。初始时设置 $P_0 = S, R_0 = I, t_0 = 0$, 迭代次数 $k = 0$, ICP 精配准算法的迭代步骤如下。

输入参数: 数据集 S 和 G ; 算法最大迭代次数 k ; 设定阈值 τ 。

第一步, 计算零姿态点云集 P_k 中每个点在 G 中的对应最近点, 生成对应点云集 Q 。

第二步, 定义最小化目标函数 e_k , 计算位姿变换参数 T_k 及配准误差 d_k 。

$$e_k = \frac{1}{N} \sum_{i=1}^N \| Q_{i,k} - P_{i,k} \|^2 \quad (20)$$

$$T_k = \begin{bmatrix} R_k & t_k \\ 0 & 1 \end{bmatrix} \quad (21)$$

$$d_k(R, t) = \frac{1}{N} \sum_{i=1}^N \| Q_{i,k} - (R_k P_{i,k} + t_k) \|^2 \quad (22)$$

第三步, 使用第 k 次估计得到的变换矩阵 T_k 更新源集: $P_{k+1} = T_k(P_0) = R_k P_0 + t_k$ 。

接下来进入迭代过程, 直到相邻两次的误差变化小于规定的阈值, 即 $d_k - d_{k+1} < \tau$, 或达到最大迭代次数时, 停止迭代。

最后输出配准结果, 求得转换矩阵 T , 其中包含旋转矩阵 R 和平移矩阵 t 。

6 实验与分析

本文算法在 Inter(R) Core(TM) i5-8250U CPU、64 位 Windows 10 操作系统上运行。开发环境为 Python 3.6, 基于 Pcl, Dlib 和 OpenCV 开发库。由于本文的头部姿态检测需要利用摄像机的内外参数进行标定及坐标系转换的计算, 因此实验数据均来自 Intel® RealSense™ Camera SR300, 简称 SR300。SR300 的景深范围为 0.2~1.2 m, 适用于室内和阳光间接照射环境, 彩色摄像头像素高达 1080p(30fps), 景深摄像头像素为 640 * 480。

为了更好地说明本文算法的精度, 我们在 x, y, z 3 个单轴上对本文算法进行实验检测。由于真实人体头部姿态很难测量, 因此我们在实验中使用仿真头部模型模拟真实人体头部的旋转, 采集头部模型俯、仰、左转、右转、左偏、右偏 6 种姿态图像, 并使用六轴运动姿态测量其头部姿态欧拉角的真实值。传感器在 x 和 z 轴的量程为 $\pm 180^\circ$, y 轴的量程为 $\pm 90^\circ$, 静态角度精度为 0.05° , 动态精度为 0.1° 。将陀螺仪佩戴在零姿态的仿真头部模型上, 然后对陀螺仪坐标系与三维头部坐标系进行标定, 从而得到在本文三维头部坐标系定义下的头部姿态的真实值。

人体头部姿态在左右偏转时具有对称性, 仰视运动范围大于俯视运动范围, 因此在 X 轴 ($-60^\circ, 60^\circ$)、 Y 轴 ($-60^\circ, 30^\circ$)、 Z 轴 ($-60^\circ, 60^\circ$) 区间内对本文算法进行实验测试, 每间隔 5° 采集一组 RGB-D 图像, 部分姿态下的人体头部 RGB-D 图像如图 7 所示。

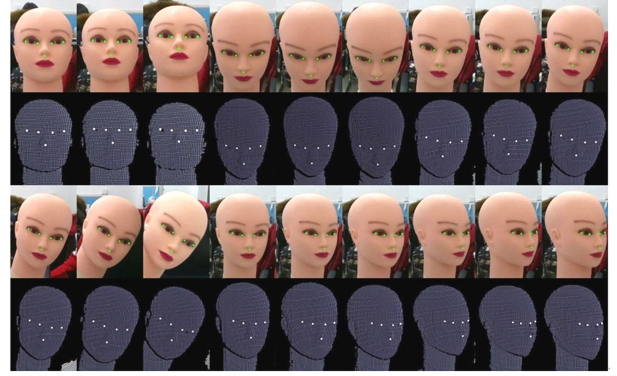


图 7 部分绕 x, y, z 轴偏转的 RGB-D 图像

假设 α, β, γ 为 x, y, z 轴上欧拉角的真实值, $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$ 为检测值, 使用平均绝对误差 (Mean Absolute Error, MAE) 作为头部姿态检测算法的评价准则。 x, y, z 单轴上头部姿态检测的绝对误差为:

$$\begin{cases} \Delta\alpha = \frac{1}{N_1} \sum_{n=1}^{N_1} (|\hat{\alpha}_n - \alpha_n|) \\ \Delta\beta = \frac{1}{N_2} \sum_{n=1}^{N_2} (|\hat{\beta}_n - \beta_n|) \\ \Delta\gamma = \frac{1}{N_3} \sum_{n=1}^{N_3} (|\hat{\gamma}_n - \gamma_n|) \end{cases} \quad (23)$$

三轴平均绝对误差为:

$$\Delta = \frac{N_1 \cdot \Delta\alpha + N_2 \cdot \Delta\beta + N_3 \cdot \Delta\gamma}{N_1 + N_2 + N_3} \quad (24)$$

图 8 为本文头部姿态粗检测算法在 x, y, z 3 个轴上对 66 个不同姿态图像进行检测的误差分析, 横轴为绕坐标轴旋转的欧拉角角度, 纵轴为绝对误差值 (以欧拉角为单位)。图中

3 条折线分别代表本文粗配准算法在 x, y, z 3 个轴上的表现,水平虚线表示三轴的平均绝对误差。

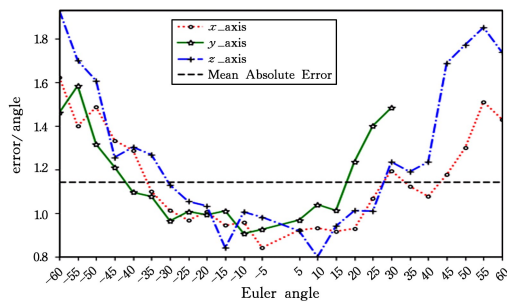


图 8 粗配准绝对误差图

由实验数据可以得出,本文提出的头部姿态粗配准算法在 3 个单轴头部正常旋转范围内的平均绝对误差为 1.141° ,在 x 轴($-60^{\circ}, 60^{\circ}$)范围内的平均误差为 1.102° ,在 y 轴($-60^{\circ}, 30^{\circ}$)范围内的平均误差为 1.089° ,在 z 轴($-60^{\circ}, 60^{\circ}$)范围内的平均误差为 1.22° 。在头部偏转角度大于 45 度时, z 轴的绝对误差高于 x 和 y 轴,这是由于头部向左右偏时,人脸特征点数据完整,没有丢失的情况。本文头部姿态检测算法在 x 轴($-40^{\circ}, 40^{\circ}$)区间的准确率和稳定性表现非常好,绝对误差均低于三轴的平均绝对误差。从图 8 可以看出,绕 z 轴旋转即左右旋转时,当欧拉角在 35 度以上时,姿态检测的绝对误差呈上升趋势,在旋转角度为 -60° 时,检测误差达到最大值 1.932° 。因此可以得出,本文提出的粗配准算法的精度依赖于特征点检测的准确率。

图 9 为本文头部姿态精配准算法在 x, y, z 3 个轴上对 66 个不同姿态图像进行检测的误差分析。精配准算法在 x, y, z 三轴的平均绝对误差分别为 $0.667^{\circ}, 0.624^{\circ}, 0.690^{\circ}$ 。点云精配准时,ICP 算法通过配准整个头部的点云数据,可以弥补在头部姿态大时由特征点丢失或检测不准造成的误差。由图 9 实验数据可知,精配准将平均绝对误差从粗配准时的 1.141° 降低到 0.663° 。在经颅磁实际应用中,欧拉角误差允许范围为 $(0^{\circ}, 2^{\circ})$,因此本文姿态检测算法的精度可达到应用标准。

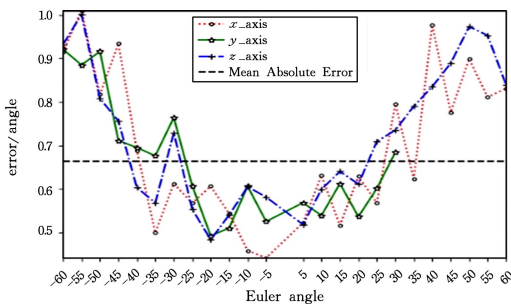


图 9 精配准绝对误差图

结束语 本文提出了一种基于 RGB-D 图像的头部姿态精确检测方法,通过深度相机获取面部特征点的三维信息,基于面部基准特征点建立三维头部坐标系,通过计算待检测三维头部坐标系与标准三维头部坐标系的旋转坐标系变换关系得到头部相对于零姿态三维标准模型的初始姿态,进一步通过 ICP 精配准得到精确的头部姿态参数。在实验中,使用仿真头部模型在 x, y, z 单轴头部正常旋转范围内欧拉角的平均绝对误差为 0.663° ;在头部偏转角度大时,误差也在应用标准范围内。

下一步将考虑基于 RGB-D 图像实现人体头部真实感的

三维重建。在头部姿态检测系统中,使用真实感头部模型,基于特征点信息实时跟踪运动中的人体头部,利用数据完整可靠的三维模型代替头部真实数据进行姿态检测,在提高检测的实时性上会有很大的突破。此外,在治疗过程中,由于不同患者头型的差异,需要磁刺激治疗的颅内神经区域位置只能通过医生经验来判断。通过建立真实感头部模型,可以得到颅内各神经区域的精确位置坐标,在诊疗过程中不仅减少医疗人员配置,而且为用户提供更加有效的治疗。

参 考 文 献

[1] MURPHY-CHUTORIAN E. Head Pose Estimation in Computer Vision: A Survey [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(4): 607-626.

[2] MEYER G P, GUPTA S, FROSIO I, et al. Robust Model-Based 3D Head Pose Estimation [C]// IEEE International Conference on Computer Vision. IEEE, 2016.

[3] MORENCY L P. 3D Constrained Local Model for rigid and non-rigid facial tracking [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012.

[4] PAPAZOV C, MARKS T K, JONES M. Real-time 3D head pose and facial landmark estimation from depth images using triangular surface patch features [C]// Computer Vision & Pattern Recognition. IEEE, 2015.

[5] TAN D J, TOMBARI F, NAVAB N. Real-Time Accurate 3D Head Tracking and Pose Estimation with Consumer RGB-D Cameras [J]. International Journal of Computer Vision, 2018, 126(2/3/4): 158-183.

[6] FANELLI G. Real Time Head Pose Estimation from Consumer Depth Cameras [C]// International Conference on Pattern Recognition. Springer-Verlag, 2011.

[7] 郭知智, 周前祥, 柳志超. 基于自适应线性回归的头部姿态计算 [J]. 计算机应用研究, 2016, 33(10): 3181-3184.

[8] SUN Y, YIN L. Automatic Pose Estimation of 3D Facial Models [C]// International Conference on Pattern Recognition. IEEE, 2008.

[9] PADELERIS P, ZABULIS X, ARGYROS A. Head pose estimation on depth data based on Particle Swarm Optimization [C]// Workshop on Human Activity Understanding from 3d Data. IEEE, 2012.

[10] WANG B, LIANG W, WANG Y, et al. Head pose estimation with combined 2D SIFT and 3D HOG features [C]// Seventh International Conference on Image & Graphics. IEEE, 2013.

[11] YANG J, WEI L, JIA Y. Face pose estimation with combined 2D and 3D HOG features [C]// International Conference on Pattern Recognition. 2012.

[12] MADRIGAL F, LERASLE F, MONIN A. 3D Head Pose Estimation Enhanced Through SURF-Based Key-Frames [C]// IEEE Winter Conference on Applications of Computer Vision. IEEE Computer Society, 2018.

[13] 唐云祁, 孙哲南, 谭铁牛. 头部姿势估计研究综述 [J]. 模式识别与人工智能, 2014, 27(3): 213-225.

[14] 闵秋莎, 刘能, 陈雅婷, 等. 基于面部特征点定位的头部姿态估计 [J]. 计算机工程, 2018, 44(6): 263-269.

[15] SEEMANN E, NICKEL K, STIEFELHAGEN R. Head pose es-

timation using stereo vision for human-robot interaction [C]// IEEE International Conference on Automatic Face & Gesture Recognition. IEEE Computer Society,2004.

[16] DERKACH D,RUIZ A,SUKNO F M,et al. Head Pose Estimation Based on 3-D Facial Landmarks Localization and Regression [C]// IEEE International Conference on Automatic Face & Gesture Recognition. IEEE,2017.

[17] IIRIS L,ESCARELA S,ANBARJAFARI G. SASE:RGB-Depth Database for Human Head Pose Estimation [C]// European Conference on Computer Vision. Springer International Publishing,2016.

[18] 李成龙. 基于 RGB-D 数据的头部姿态估计研究[D]. 济南:山东大学,2017.

[19] 尹婕. 基于图像信息的点云优化研究[D]. 成都:电子科技大学,2017.

[20] ROUTRAY S,RAY A K,MISHRA C. Image denoising by preserving geometric components based on weighted bilateral filter and curvelet transform [J]. Optik,2018;S0030402618301153.

[21] 张鸿宇,刘威,许炜,等. 基于深度图像的多学习者姿态识别[J]. 计算机科学,2015,42(9):299-302.

[22] 郭连朋,陈向宁,刘彬. Kinect 传感器的彩色和深度相机标定[J]. 中国图象图形学报,2014,19(11):1584-1590.

[23] ÇELIKTUTAN O,ULUKAYA S,SANKUR B. A comparative study of face landmarking techniques [J]. Eurasip Journal on Image & Video Processing,2013,2013(1):13.

[24] MARACI M A,NAPOLITANO R,PAPAGEORGHIU A , et al. Object classification in an ultrasound video using LP-SIFT features [M]// Medical Computer Vision: Algorithms for Big Data. Springer International Publishing,2014.

[25] KAZEMI V,SULLIVAN J. One Millisecond Face Alignment with an Ensemble of Regression Trees [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society,2014.

[26] 蔡自兴,谢斌. 机器人学[M]. 北京:清华大学出版社,2015:20.

[27] 李良福,邹彬,周国良,等. 基于优化估计的深度图像修复与误差补偿方法研究[J]. 应用光学,2018(1):45-50.

[28] BESL P J,MCKAY N D. A Method for Registration of 3-D Shapes [M]. IEEE Computer Society,1992.

(上接第 333 页)

[6] BERGEN J B. Hierarchical Model-Based Motion Estimation [C]//Proceedings of Second European Conference on Computer Vision,1992,11(5):237-252.

[7] BONNIER D,LUTZ Y,LABRANCHE B. Modeling of an active TV system for surveillance operations[C]//Aerosense. International Society for Optics and Photonics,1999.

[8] JOSHI N,SZELISKI R,KRIEGMAN D J. PSF Estimation using Sharp Edge Prediction[C]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA;IEEE,2008.

[9] XU L,JIA J. Two-Phase Kernel Estimation for Robust Motion Deblurring[J]. Proceedings of European Conference on Computer Vision,2010,6311(9):157-170.

[10] CHO S,LEE S. Fast motion deblurring[C]// ACM Transactions on Grphics,2009,28(5):1-8.

[11] 方帅,赵育坤,李心科,等. 基于光照估计的夜晚图像去雾[J]. 电子学报,2016,44(11):2569-2575.

[12] 王一宁,秦品乐,李传朋,等. 基于残差神经网络的图像超分辨率改进算法[J]. 计算机应用,2018,38(1):246-254.

[13] TANG K,YANG J,WANG J. Investigating Haze-Relevant Features in a Learning Framework for Image Dehazing[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2014.

[14] CAI B,XU X,JIA K, et al. DehazeNet: An End-to-End System for Single Image Haze Removal [J]. IEEE Transactions on Image Processing,2016,25(11):5187-5198.

[15] 梁中豪,彭德巍,金彦旭,等. 基于交通场景区域增强的单幅图像去雾方法[J]. 计算机应用,2018,38(5):204-210.

[16] 隋阳,孟钊楠,董玮,等. 基于径向基函数神经网络直接提取布里渊散射谱温度的方法[J]. 光学学报,2018,38(12):394-400.

[17] 张育贵,王义,杨人静. 基于径向基神经网络的天气预测模型[J]. 贵州大学学报(自然版),2018,35(1):69-72.

[18] 熊聪聪,潘璇,赵奇,等. 多模式集成的 RBF 神经网络天气预报[J]. 天津科技大学学报,2014,29(1):75-78.

[19] 朱福珍,李金宗,朱兵,等. 基于径向基函数神经网络的超分辨率图像重建[J]. 光学精密工程,2010,18(6):1444-1451.

[20] ABBASCHIAN L,LIMA M S F D. Cracking susceptibility of aluminum alloys during laser welding[J]. Materials Research,2003,6(2):273-278.

[21] 张健,杨锐. 基于径向基函数神经网络的激光焊接熔池光强分布预测[J]. 中国激光,2010,37(7):1856-1860.

[22] GRABEC I. Extraction of physical laws from joint experimental data[J]. The European Physical Journal B - Condensed Matter and Complex Systems,2005,48(2):279-289.

[23] YOU Y L,KAVEH M. Blind image restoration by anisotropic regularization[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society,1999,8(3):396-407.

[24] 胡海蕾. LED 照明光学系统的设计及其阵列光照度分布研究[D]. 福州:福建师范大学,2005.

[25] 熊玲玲,吕百达. 描述激光二极管远场光强分布的理论模型[J]. 强激光与粒子束,2008,20(2):201-206.

[26] CHEN Y,YANG L,ZENG Z, et al. Degradation in LED night vision imaging and recovery algorithms[J]. Optik - International Journal for Light and Electron Optics,2017,144(144):240-245.

[27] KATSAGGELOS A K. Digital image restoration [J]. Proceedings of the IEEE,1978,66(7):812-812.

[28] YANG F,CHOI W,LIN Y. Exploit All the Layers:Fast and Accurate CNN Object Detector with Scale Dependent Pooling and Cascaded Rejection Classifiers[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society,2016.

[29] HAN H,ZHANG X,GE W. Performance Evaluation of Underwater Range-gated Viewing Based on Image Quality Metric [C]//International Conference on Electronic Measurement & Instruments 2009(ICEMI 09). Beijing,2009:441-444.