

基于深度相机的水平视角估计及校准方法

倪剑帆,汪建明,周小芹,刘小峰

(河海大学 物联网工程学院 常州市特种机器人与智能技术重点实验室,江苏 常州 213022)

摘要:提出一种基于 Kinect 深度相机水平视角估计及校准方法。用 Kinect 捕捉人的脸部和估计头部姿势,采用基于图像梯度信息的方法从 2 维脸部图像定位瞳孔中心点,利用 Kinect 固有参数,得到三维坐标系中的瞳孔中心点。结合预设眼球中心点,可以很方便、稳定地估算出水平视角。该方法采用预定义的眼球中心点经验值,无需单独校准每个人的眼球中心点,提高了该视线检测系统的普遍适用性。为了提高视线角度的精确性,设计了一套能够呈现注视目标并可采集注视角度的实验系统。应用该实验系统采集数名实验者视角估计值和真实数据,然后建立线性回归模型。通过拟合函数校准估计的视角,进一步提高了视角精度。实验结果表明,对于不同的人,该方法都能快速而准确地估计出视角。

关键词:校准方法;水平视线角度估计;Kinect 深度相机;线性回归模型;拟合函数

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2017)01-0039-05

doi:10.3969/j.issn.1673-629X.2017.01.009

An Approach of Horizontal Gaze Angle Estimation and Calibration Using Depth Camera

NI Jian-fan, WANG Jian-ming, ZHOU Xiao-qin, LIU Xiao-feng

(Changzhou Key Laboratory of Robot and Intelligent Technology, College of IoT Engineering,
Hohai University, Changzhou 213022, China)

Abstract: An approach to estimate and calibrate the horizontal gaze angle using the Kinect depth camera is proposed. First of all, the Kinect is used to capture the face image and estimate the head pose, localizing the 2D pupil center by means of image gradient. Then the 3D pupil center is obtained by the intrinsic parameters of the Kinect. With the pupil center and an empirical eyeball center value, the horizontal gaze angle can be estimated conveniently. Unlike the previous methods which need to calibrate the eyeball center for every person, the method is based on the predefined empirical eyeball center, which the gaze estimation system is robust and convenient. In order to improve the accuracy of the gaze angle, an experimental system is built which can collect the gaze angle and display the gaze target. The several participants' data of measurement data and the ground truth are collected. A linear regression model is built from the errors between the ground truth and the measured data, and used to correct the accuracy of the estimated gaze direction. The experiment shows that the approach can fast and accurately estimate the gaze angle.

Key words: calibration approach; horizontal gaze angle estimation; the Kinect depth camera; linear regression model; fitting function

0 引言

眼神交流在日常生活中有着重要的作用。相应的视线检测就成为了一个热门的研究课题,它在人机交互、残疾人交流、安全监控和疲劳驾驶中都有着广泛的应用。

通常眼睛视线检测的方法有两类:基于外表的方法和基于特征/模型的方法。前者^[1-3]是通过高维的眼部图像直接映射到低维空间的视线坐标上。文献[2]保留了高维向量中更有代表性的特征。文献[3]

提出了一种线性回归方法,在低维空间中实现精确的视线估计,但是这种方式容易受到光线等的干扰。后者一般是根据头部和眼部的特征来建立 3D 模型。文献[4-5]通过建立 3D 眼睛模型来检测视线,通过眼睛模型得到眼部特征点,然后通过获得眼球中心点和瞳孔中心点的 3 维坐标来得到视线角度。文献[6]用 3D 模型估计头部姿势和视线的方向向量,这种基于模型的方法较简单且鲁棒性较好,但是其结果的好坏取决于头部估计和瞳孔追踪的精确性,而文献中也没有相

收稿日期:2016-03-10

修回日期:2016-06-14

网络出版时间:2017-01-04

基金项目:中央高校基本科研业务费专项资金(2015B38214);常州市特种机器人与智能技术重点实验室开放基金项目(CZSR2014003)

作者简介:倪剑帆(1991-),男,硕士研究生,研究方向为人机交互、视线检测;刘小峰,教授,研究方向为人机交互、智能传感与计算。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170104.1028.050.html>

应的校准方法。

早期的视线检测技术基本都采用普通摄像头采集数据^[7],对于视线的估计都是在 2 维空间,一般只能检测视线方向而无法得到具体的视线角度。近年,微软开发的 Kinect 传感器被广泛用于视线检测,它不需要额外的硬件设备,并且在检测视线时头部可以自由转动。文献[1,5,8-9]均使用 Kinect 直接得到视线目标位置,从而得到真实值。由于 Kinect 本身对物体位置的估计不够精确,该方法会对结果产生一定的误差。文献[5,10]通过已知坐标点的目标物来校准眼球中心点。虽然该方法得到的结果较为准确,但是需要对每个人的眼球中心点进行单独校准。

文中提出一种基于 Kinect 的水平视角估计及校准方案。通过 Kinect 捕捉脸部图像,并估计头部姿势。通过文献[5]中提出的简化 3D 眼睛模型建立眼部特征点间关系。采用基于梯度信息的瞳孔中心定位算法^[11]检测瞳孔。为了避免对每个人都进行眼球中心点校准,采用了眼球中心点经验值。眼睛的视线角度可通过 Kinect 坐标系中的眼球中心点和瞳孔中心点得到。为了提高视线角度的精度,设计了一套实验系统校准视线估计误差,它能够采集实验者视角的测量值和真实值数据。在实验中,采用 Kinect 获得视角测量值,通过定点的实际坐标得到真实值。鉴于实验中存在的一些不足,改进了原先的实验系统。通过得到的数据建立线性回归模型,进一步提高了视角精度,而且该方法适用性好,操作简便,精度高。

1 算法描述

文中采用 Kinect 传感器估计头部姿势,并采用基于图像梯度信息的瞳孔中心定位方法^[11]和简化的 3D 眼球模型^[5]检测视线角度。

1.1 瞳孔中心定位

不同于文献[12]中需佩戴一套设备来检测瞳孔位置,文中使用文献[11]中基于梯度信息的 2 维瞳孔定位方法,避免了检测设备和人皮肤的直接接触,使得测试者可以自由灵活地转动头部。通过单位化的位移向量和梯度向量的数量积得到瞳孔中心。具体的最优瞳孔中心 p 由如式(1)所得:

$$p = \operatorname{argmax}_{p^*} \left\{ \frac{1}{n} \sum_{i=1}^n (d_i^T g_i)^2 \right\} \quad (1)$$

其中, p^* 是随机选定的瞳孔中心; g_i 是在任意位置 q_i 处的梯度向量; d_i 是 q_i 和 p^* 之间的单位化位移向量,其方向和 g_i 一致。

最终得到瞳孔中心坐标 $p = (x_p, y_p)$ 。然后用 Kinect 把 2D 坐标转换到 3D 坐标,如式(2)所示:

$$\frac{x}{x_p - c_{\text{center}}} = -\frac{y}{y_p - r_{\text{center}}} = \frac{z}{\lambda} \quad (2)$$

其中, $P = (x, y, z)$ 为瞳孔 3D 坐标; $(c_{\text{center}}, r_{\text{center}})$ 为图像的中心坐标; λ 为 Kinect 相机焦距。 c_{center} , r_{center} 和 λ 都是 Kinect 相机的内部参数。

1.2 眼部视线估计

文献[13]中提出了比较完整的眼球模型和视线关系的原理图,而文中采用文献[5]提出的简化 3D 眼球模型,不过文献[5]头部坐标系中眼球中心点要对每个人进行单独校准。文中采用了一个经验值:在头部坐标系中,左眼球中心点坐标 $C_{\text{left}} = (-0.033 \ 8, 0.038 \ 5, -0.139 \ 8)^T$;右眼球中心点坐标 $C_{\text{right}} = (0.033 \ 1, 0.038 \ 5, -0.139 \ 8)^T$ 。根据文献[13]可知,眼球中心到瞳孔中心的距离,眼球中心到角膜中心 C_p 的距离,光轴(optical axis) $\overrightarrow{C_p P}$ 和视轴(visual axis) $\overrightarrow{C_p M}$ 之间的角度 θ 均为常量。然后通过眼球模型中的相应关系进行分析与计算。

文献[14]中用头部几何模型和估计算法对头部姿势进行估计,而文中 Kinect 可以直接估计人头部的偏转姿势而不需要其他辅助设备,它能得到头部的位移矩阵 T 和旋转矩阵 R 。把头部坐标系中的眼球中心点 C_{head} 转换为 Kinect 坐标系中的 C 。

$$C = R * C_{\text{head}} + T \quad (3)$$

视轴(visual axis) $\overrightarrow{C_p M} = (x_{cm}, y_{cm}, z_{cm})$ 可以通过已知的常量关系得到,如式(4):

$$\cos \theta = \frac{\overrightarrow{C_p M} \cdot \overrightarrow{C_p P}}{\|\overrightarrow{C_p M}\| \|\overrightarrow{C_p P}\|} \quad (4)$$

$$\alpha = \arctan(x_{cm}/z_{cm}) \quad (5)$$

其中, M 为视线目标。

最终得到水平视线角度 α (测量值)。对于垂直视线角度,文中不进行讨论。

2 校准方法

为了提高视角精度,提出一种通过实验对视角进行校准的方法。在实验中由上述算法直接得到测量值,通过实际固定点的坐标得到真实值。实验系统通过采集几名视力正常测试者的测量值和真实值,建立线性回归模型,得到不同情况下的测量值和真实值的拟合曲线,以校准视角。鉴于实验系统存在的一些不足,对原始的实验系统进行改进。

2.1 原始实验系统

文中设计的原始实验系统如图 1 所示。其中,1 代表激光笔;2 代表 Kinect;3 代表可移动黑板。

如图 1 所示,在黑板上与人视线齐平的是一横排视线点,与激光笔齐平的是一横排激光点。实验中,测

试者和 Kinect 距离约为 70 cm,确保能捕捉到头部图像。测试者需要后背贴墙坐直,视线平视正前方,将黑板上看到的首个点标记为 0 点,也称为固定点。然后分别在它的水平左右两边取若干点,通过保持眼睛直视前方的情况下分别向左右转动头部,使得视线恰好到所标点处,在转动过程中保持头部中心点到黑板的距离不变。为了便于视线恰好到达所标点处,采用激光笔校准头部偏转,将激光笔固定在实验者戴的帽子上,使激光笔在头部正中心,且与实验者的视线保持水平。



图1 原始实验系统

表1 原实验视角误差的均值(mean)和标准差(std)

No.	左眼 (mean±std)		右眼 (mean±std)	
	向左看	向右看	向左看	向右看
1	-4.283±3.168	2.404±2.723	2.23±2.331	3.872±2.382
2	-3.733±2.927	-0.749±3.627	-1.113±2.159	1.676±2.052
3	-3.306±2.774	0.805±2.391	-0.436±2.102	2.483±1.939
4	-3.39±2.549	0.805±2.357	-0.436±2.063	2.483±1.922
5	-2.218±2.135	1.872±2.267	0.555±2.248	3.059±1.97
6	-2.845±2.268	0.841±1.649	1.172±1.819	3.302±2.249
7	-3.397±2.219	-1.066±2.076	-0.266±2.18	1.797±2.156
8	-3.405±2.079	-1.344±2.322	0.645±1.85	2.235±1.76
9	-3.667±1.919	-1.641±2.035	-0.652±2.329	2.724±1.996
总的	-3.329±2.544	0.256±2.802	0.313±2.337	2.621±2.173

2.2 线性回归模型

线性回归模型是一种用数理统计中的回归分析来确定两个或多个变量间的定量关系的统计分析方法,它能够通过现有变量之间的关系得到它们之间的预测模型。文中采用最小二乘法拟合曲线,考察 9 个人的数据在不同情况下误差和测量值的关系,如图 2 所示。

从中可以看到,在四种情况下它们的误差有着明显的不同。所以用线性回归模型对四种情况分别进行拟合,得到:

$$y_{lelw}(x_i) = 0.820\ 4x_i - 0.065\ 5 \tag{7}$$
$$y_{lerw}(x_i) = 0.888x_i + 1.891\ 2 \tag{8}$$
$$y_{relw}(x_i) = 0.909x_i - 0.907 \tag{9}$$

实验中的真实值 φ 可通过式(6)得到。

$$\varphi = \arctan\left(\frac{\text{dist}(0i)}{\text{dist}(OP)}\right) \tag{6}$$

其中, $\text{dist}(OP)$ 是眼睛到黑板的距离; $\text{dist}(0i)$ 是 0 点到 i 点的距离。

通过移动黑板到实验者的距离来改变真实值 φ 的数值,该方法比直接采用 Kinect 得到目标的真实值更准确。测量值 α 可以通过算法直接获得。通过 9 位视力正常测试者的 780 组实验数据对该方法进行结果分析,实验时对每个人都要调整视线点和激光点的位置以及 Kinect 的高度。其中,真实值 φ 和测量值 α 误差的均值和标准差如表 1 所示。

表 1 为 9 位测试者真实值和测量值误差的均值和标准差,从中可看出每个人之间的均值和标准差没有明显的区别。当左眼向左看时,测量值基本比真实值大;当右眼向右看时,测量值基本比真实值小;当左眼向右看和右眼向左看时,真实值和测量值误差要小于另外两种情况,但是它们误差的稳定性和其他两种情况下相差不多,甚至比其他两种情况的波动还要大一些。

$$y_{rerw}(x_i) = 1.130\ 6x_i + 0.938\ 3 \tag{10}$$

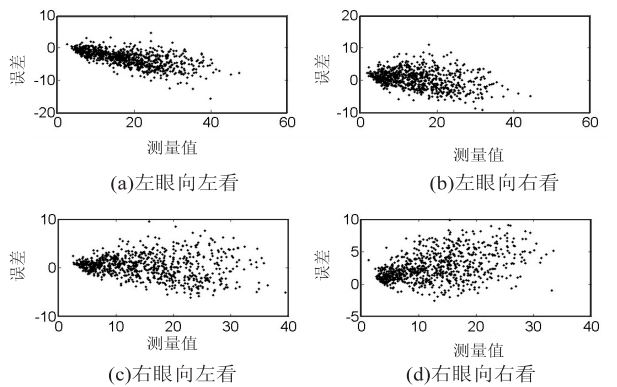


图2 四种情况下的误差和测量值

其中, $y_{lelw}, y_{lerw}, y_{relw}, y_{rerw}$ 分别表示左眼向左看,左

眼向右看,右眼向左看,右眼向右看校准后的角度; x_i 表示测量值。

然后根据得到的校准后角度和真实值的误差再次进行比较,可得经过校准后误差比之前更稳定,不过测量值与真实值之间的误差还在 10° 以内。

2.3 改进后实验系统

考虑到原实验系统存在的一些不足,比如随着头部转动幅度的变大,Kinect 对头部的估计误差会增大;头部转动后人眼到定点的距离也会发生微小变化,导致真实值不精确。改进后的实验系统如图 3 所示。其中,1 代表头部支架;2 代表 Kinect;3 代表显示屏。

Kinect 与测试者的距离还是保持 70 cm。实验者头部固定在支架上,保持头部不动,通过转动眼睛来注视屏幕上的点,这样 Kinect 对头部的估计和计算得到的真实值都会变得更精确。实验中通过变化屏幕与实验者的距离来改变真实值。收集、处理数据同原始实验系统。

改进后的实验只是改变了注视目标的方式。不再是通过转动头部,保持眼睛正视前方来检测视线,而是

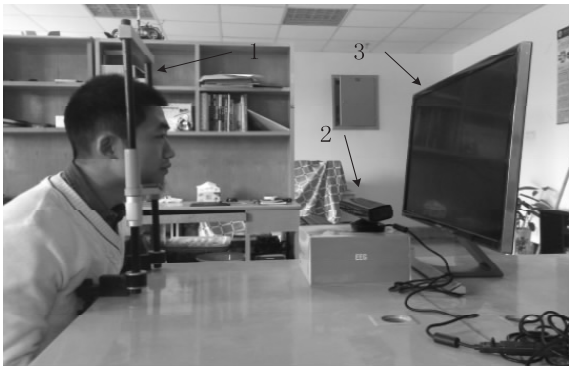


图 3 改进后实验系统

通过保持头部不动,转动眼睛来检测注视角度。这样就能避免一些误差。实验采集了 10 名测试者总共 300 组测量值和真实值数据,具体的误差的均值和标准差如表 2 所示。

与表 1 相比,表 2 中的数据更稳定,误差更小。当看向同一个方向,导致真实值和测量值误差不同的原因可能是人看事物的习惯。从表 2 可知,当实验者向左看,右眼比左眼转动的幅度大,导致右眼的测量值大于真实值。当看向右边时,同样也存在这个问题。

表 2 改进后实验视角误差的均值 (mean) 和标准差 (std)

No.	左眼 (mean±std)		右眼 (mean±std)	
	向左看	向右看	向左看	向右看
1	1.014±1.674	-1.906±2.876	1.292±2.413	3.082±1.821
2	3.098±2.213	-4.092±2.114	1.948±2.672	0.097±2.09
3	2.442±1.864	-2.579±2.112	0.847±0.853	0.199±1.496
4	0.957±2.415	-1.934±2.641	0.609±1.4	0.238±2.738
5	2.081±2.276	-2.579±2.112	-0.376±1.415	0.613±2.117
6	1.388±1.336	-1.818±1.646	-0.354±1.604	2.419±2.006
7	-0.334±1.687	-1.887±2.082	-2.271±1.828	0.721±2.099
8	0.994±2.012	0.752±1.545	-2.084±2.277	3.107±2.465
9	1.144±1.397	-0.978±1.725	-0.397±1.316	2.18±1.279
10	1.592±1.216	-0.57±2.028	-0.499±1.124	2.331±2.539
总的	1.438±2.058	-1.758±2.407	-0.129±2.198	1.499±2.413

然后,通过线性回归模型得到了四种不同情况下的校准函数。

$y_{lelw}(x_i) = 1.021\ 2x_i + 1.132\ 8$ (11)

$y_{lerw}(x_i) = 0.849\ 3x_i + 0.877\ 6$ (12)

$y_{relw}(x_i) = 0.907\ 5x_i + 1.347\ 6$ (13)

$y_{rerw}(x_i) = 0.973\ 9x_i + 1.870\ 7$ (14)

与式(7)~(10)一样,通过得到测量值 x ,就能得到相应校准后的值 y 。改进实验后,得到的测量值更接近真实值,它们之间的误差基本可以控制在 5° 以内。

两个实验经过校准后具体误差的均值和标准差由表 3 所示。通过比较表 1 和表 2 中校准前数据可知,改进后实验系统校准后误差波动更小。

经过校准函数后,误差的均值明显减小,而且误差变得更稳定。由于改进后实验本来的误差就小,所以校准后的效果没有原实验那么明显,不过改进后的实验校准后误差波动更小。

表 3 原始实验系统和改进后实验系统校准后误差的均值和标准差 (mean(e-6) ±std)

系统	左眼		右眼	
	向左看	向右看	向左看	向右看
原实验	-1.4±1.97	0.3±2.66	-0.4±2.33	-0.8±1.99
改进后实验	1.3±2.04	-1.0±2.03	1.2±2.07	3.7±2.40

3 结束语

文中提出一种基于深度相机的水平视角估计及校准方法。首先采用图像的梯度法在 2 维图像中找到瞳孔中心点,并用 Kinect 估计头部姿势,然后把瞳孔坐标转换到 Kinect 3 维坐标系中。结合预定义的眼球中心点经验值,通过简化的 3D 眼球模型,得到在 Kinect 坐标系中的水平视线角度。为了提高精度,设计了一套实验系统进行校准。实验中,通过实际定点的坐标得到真实值,通过文中算法直接得到测量值。然后建立线性回归模型,采用拟合函数校准得到测量视角。由最终得到的误差均值和标准差的数据说明,该方法适合所有人,并且能够方便有效地校准视线角度,鲁棒性好。

为了将这套视线检测系统应用于机器人控制上,使用了经验值眼球中心点,这样就能适用于所有人。当然对于瞳孔点定位的准确性和实时性还有待提高。

参考文献:

[1] Mora K A F, Odobez J M. Person independent 3d gaze estimation from remote rgb-d cameras[C]//20th IEEE international conference on image processing. Melbourne, Australia: IEEE, 2013;2787-2791.

[2] Tan K H, Kriegman D J, Ahuja N. Appearance-based eye gaze estimation[C]//Sixth IEEE workshop on applications of computer vision. Washington DC, USA: IEEE, 2002;191-195.

[3] Lu F, Sugano Y, Okabe T, et al. Inferring human gaze from appearance via adaptive linear regression[C]//IEEE international conference computer vision. Barcelona, Spain: IEEE, 2011;153-160.

[4] Chen J, Ji Q. 3D gaze estimation with a single camera without IR illumination[C]//19th international conference on pattern

recognition. Florida, USA: IEEE, 2008;1-4.

[5] Li J, Li S. Eye-model-based gaze estimation by rgb-d camera[C]//IEEE conference on computer vision and pattern recognition workshops. [s. l.]: IEEE, 2014;606-610.

[6] Heinzmann J, Zelinsky A. 3-D facial pose and gaze point estimation using a robust real-time tracking paradigm[C]//Third IEEE international conference on automatic face and gesture recognition. Nara, Japan: IEEE, 1998;142-147.

[7] 李 涛. 基于视线角度的人眼视线检测研究[J]. 计算机技术与发展, 2009, 19(8): 37-40.

[8] Jafari R, Ziou D. Gaze estimation using Kinect/PTZ camera[C]//IEEE international symposium on robotic and sensors environments. Magdeburg, Germany: IEEE, 2012;13-18.

[9] Mora K A F, Odobez J M. Gaze estimation from multimodal Kinect data[C]//IEEE computer society conference on computer vision and pattern recognition workshops. Rhode Island, USA: IEEE, 2012;25-30.

[10] Xiong X, Cai Q, Liu Z, et al. Eye gaze tracking using an RGBD camera: a comparison with a RGB solution[C]//Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. Seattle: ACM, 2014;1113-1121.

[11] Timm F, Barth E. Accurate eye centre localisation by means of gradients[C]//VISAPP. Algarve, Portugal: [s. n.], 2011;125-130.

[12] 沈海平, 冯华君, 徐之海. 基于瞳孔检测的注视点跟踪系统[J]. 光电子·激光, 2005, 16(8): 961-964.

[13] Guestrin E D, Eizenman M. General theory of remote gaze estimation using the pupil center and corneal reflections[J]. IEEE Transactions on Biomedical Engineering, 2006, 53(6): 1124-1133.

[14] 于 琼. 视频图像序列内的视线跟踪研究[D]. 武汉: 武汉理工大学, 2011.