



基于表观的归一化坐标系分类视线估计方法

戴忠东¹,任敏华²

(1.上海复控华龙微系统技术有限公司,上海 200439; 2.中国电子科技集团公司第三十二研究所,上海 201808)

摘要: 视线估计能够反映人的关注焦点,对理解人类的情感、兴趣等主观意识有重要作用。但目前用于视线估计的单目眼睛图像容易因头部姿态的变化而失真,导致视线估计的准确性下降。提出一种新型分类视线估计方法,利用三维人脸模型与单目相机的内在参数,通过人脸的眼睛与嘴巴中心的三维坐标形成头部姿态坐标系,从而合成相机坐标系与头部姿态坐标系,并建立归一化坐标系,实现相机坐标系的校正。复原并放大归一化得到的灰度眼部图像,建立基于表观的卷积神经网络模型分类方法以估计视线方向,并利用黄金分割法优化搜索,进一步降低误差。在MPIIGaze数据集上的实验结果表明,相比已公开的同类算法,该方法能降低约7.4%的平均角度误差。

关键词: 视线估计;单目眼睛图像;头部姿态;归一化坐标系;黄金分割法;卷积神经网络

开放科学(资源服务)标志码(OSID):



中文引用格式:戴忠东,任敏华.基于表观的归一化坐标系分类视线估计方法[J].计算机工程,2022,48(2):230-236.

英文引用格式:DAI Z H,REN M H.Gaze estimation method using normalized coordinate system classification based on apparent[J].Computer Engineering,2022,48(2):230-236.

Gaze Estimation Method Using Normalized Coordinate System Classification Based on Apparent

DAI Zhongdong¹,REN Minhua²

(1.Shanghai Fudan-Holding Hualong Microsystem Technology Co.,Ltd.,Shanghai 200439,China;

2.The 32nd Research Institute of China Electronics Technology Group Corporation,Shanghai 201808,China)

[Abstract] Gaze estimation can naturally reflect people's focus of attention, and plays an important role in understanding human emotions, interests and other subjective consciousness. However, monocular images used for gaze estimation tend to be distorted due to head pose changes, which reduces the accuracy of gaze estimation. This paper proposes a new classification-based gaze estimation method. A three-dimensional face model and the inherent parameters of monocular camera are used to form a head pose coordinate system through the three-dimensional coordinates of the center of eye and mouth. The camera coordinate system and the head pose coordinate system are combined to establish a normalized coordinate system, and the camera coordinate system is corrected. The gray eye image is obtained by restoration, magnification and normalization. Finally a classification method using an appearance-based convolution neural network model is established to estimate gaze direction, and the golden section method is used to optimize the search process and further reduce errors. The experimental results show that compared with other similar methods, the proposed method can reduce the average angle errors by about 7.4% on the commonly used MPIIGaze test dataset.

[Key words] gaze estimation; monocular eye image; head pose; normalized coordinate system; golden section method; Convolutional Neural Network(CNN)

DOI:10.19678/j.issn.1000-3428.0059684

0 概述

近年来,视线估计在人机交互中得到广泛应用,例如使计算机理解人类感兴趣的区域,借此进行性格分析、商品推荐、游戏娱乐等。目前,视线估计大多使用多目图像^[1-3],但实际场景相对复杂,该方法

使用效果不佳。另外,还有研究关注单目图像视线估计^[4-6]。然而,嵌入式应用时图像分辨率较低,且头部姿态变化存在一定的随机性,因此视线估计的准确性依然是目前面临的最大挑战。

单目摄像视线估计可分为2大类:基于形状的视线估计方法和基于表观的视线估计方法^[7]。基于

基金项目:上海市国防科技工业办公室支持基金(GFKJ-2019-060)。

作者简介:戴忠东(1968—),男,高级工程师、硕士,主研方向为计算机视觉、人工智能;任敏华,研究员、硕士。

收稿日期:2020-10-10 修回日期:2021-02-07 E-mail:daizhongdong@fkh.sh.cn

形状的方法依赖于为眼睛、瞳孔和周围区域建立特定模型以计算瞳孔中心的位置,从而进行视线估计。基于表观的方法通常从眼睛区域中提取多维特征进行视线估计。基于形状的方法稳定性较差,模型经常过于理想化。基于表观的方法使用回归计算视线方向,虽然平均角度误差仍需进一步降低,但其在图像光照差、分辨率低等条件下具有较好的稳定性,具备较大的提升空间,在视线估计研究中具有重要作用。

在基于表观的方法的研究中,SUGANO等^[8]使用大量的交叉研究对象数据来训练三维视线模型。为保持一致性,收集了最大的、完全校准的多视角视线数据集来进行三维重建,并使用包括随机森林^[9]、k-NN^[10]、自适应线性回归^[11]等方法从眼睛图像中提取视线方向。为得到更好的结果,SUGANO等^[12]构建了一个更大的数据集 MPIIGaze,其中包含 15 个参与者,并提出一种利用卷积神经网络(Convolutional Neural Network, CNN)模型进行基于表观的视线估计方法,对归一化后的眼睛图像和头部姿态角进行训练。SUGANO等^[13]对神经网络进行改进,并采用更深层的模型,获得了更好的性能。与早期使用头部姿态信息作为卷积神经网络模型部分输入不同,RANJAN等^[14]将数据集汇聚为头部姿态的不同组别,形成具有分支结构的卷积神经网络模型,提高了视线估计对不同头部姿态的稳定性。ZHANG等^[15]则使用空间权值卷积神经网络模型估计视线方向,空间权值卷积神经网络模型采用全脸图像作为输入,不经过归一化,适合各种头部姿态,其特点在于空间权重更关注眼睛区域。但上述基于表观的方法存在一些不足。文献[8, 12-13]中介绍的归一化方法不能建立相机坐标系,因此使用了一个包含所有 15 名参与者的平均面部模型来定位三维空间中的眼睛中心。因此,归一化图像的眼睛中心和实际视线方向相比并不精确。此外,上述方法还使用了回归方法来获得预测的视线方向,但回归易受带噪点的低分辨率图像以及归一化方法的影响。

本文提出采用归一化坐标系方法来校准相机坐标系,以复原人眼图像。同时,建立基于表观的卷积神经网络模型,分类方法估计视线方向,并利用黄金分割法进行优化进一步降低误差。

1 基于表观的归一化坐标系分类视线估计

1.1 框架基础

本文所提方法主要包括 2 个阶段:分类视线估计及黄金分割法寻优。其中分类视线估计如图 1 所示,通过对归一化眼睛图像的多分类实现。而黄金分割法则对分类数进行搜索以进一步优化。图 2 是

本文提出的归一化坐标系方法所形成的眼睛图像,虽然原始图像对应了不同人、不同头部姿态与光照,但可以看到人眼图像形状基本一致。

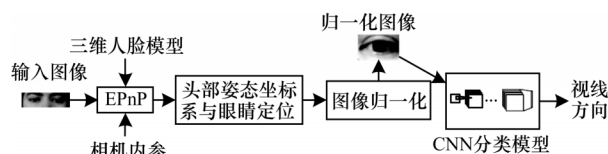


图1 基于表观的归一化坐标系分类视线估计

Fig.1 Gaze estimation based on apparent normalized coordinate system classification

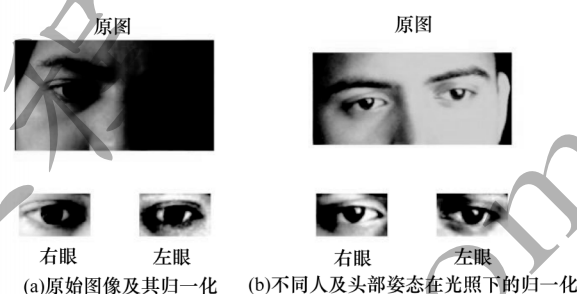


图2 不同的人眼图像及其归一化

Fig.2 Different eyes photo and normalization

在视线估计中,使用注视点的视线向量及其相关的失配角进行描述或评价。注视点是注视屏幕与视线的交汇点,可以通过屏幕平面的二维坐标或相机坐标系中的三维坐标来表示。在图 3 中,视线向量(e_x, e_y, e_z)是相机坐标系中从眼睛中心指向屏幕平面的注视点单位向量。视线向量在垂直和水平方向上的投影角分别称为垂直角 θ 和水平角 φ 。垂直角和水平角计算式如下:

$$\begin{cases} \theta = \arcsin(e_y) \\ \varphi = \arctan(e_x/e_z) \end{cases} \quad (1)$$

式(1)也可以表示为:

$$\begin{cases} e_x = \cos(\theta) \sin(\varphi) \\ e_y = \sin(\theta) \\ e_z = \cos(\theta) \cos(\varphi) \end{cases} \quad (2)$$

实际视线向量(e_{x1}, e_{y1}, e_{z1})和预测视线向量(e_{x2}, e_{y2}, e_{z2})间的失配角 $\Delta\alpha$ 被用来评价视线估计的性能,失配角被定义为:

$$\Delta\alpha = \arccos(e_{x1} \times e_{x2} + e_{y1} \times e_{y2} + e_{z1} \times e_{z2}) \quad (3)$$

实际视线向量和预测视线向量的失配角也可以通过它们在垂直和水平方向上的投影来计算,例如:

$$\Delta\alpha = \arccos(e_{x1} \times e_{x2} + e_{y1} \times e_{y2} + e_{z1} \times e_{z2}) \approx \arccos(\cos(\Delta\theta) \cos(\Delta\varphi)) \quad (4)$$

证明如下:

$$\begin{aligned}
\Delta\alpha &= \arccos(\mathbf{e}_{x1} \times \mathbf{e}_{x2} + \mathbf{e}_{y1} \times \mathbf{e}_{y2} + \mathbf{e}_{z1} \times \mathbf{e}_{z2}) = \\
&\arccos(\cos(\theta_1)\cos(\theta_2)\sin(\varphi_1)\sin(\varphi_2) + \\
&\sin(\theta_1)\sin(\theta_2) + \cos(\theta_1)\cos(\theta_2)\cos(\varphi_1) \\
&\cos(\varphi_2)) = \\
&\arccos(\cos(\theta_1)\cos(\theta_2)(\cos(\varphi_1)\cos(\varphi_2) + \\
&\sin(\varphi_1)\sin(\varphi_2)) + \sin(\theta_1)\sin(\theta_2)) = \\
&\arccos(\cos(\theta_1)\cos(\theta_2)\cos(\varphi_1 - \varphi_2) + \\
&\sin(\theta_1)\sin(\theta_2)) \approx \\
&\arccos(\cos(\theta_1)\cos(\theta_2)\cos(\varphi_1 - \varphi_2) + \\
&\sin(\theta_1)\sin(\theta_2)\cos(\varphi_1 - \varphi_2)) = \\
&\arccos(\cos(\theta_1)\cos(\theta_2) + \\
&\sin(\theta_1)\sin(\theta_2))\cos(\varphi_1 - \varphi_2) = \\
&\arccos(\cos(\theta_1 - \theta_2)\cos(\varphi_1 - \varphi_2)) \\
&\arccos(\cos(\Delta\theta)\cos(\Delta\varphi))
\end{aligned}$$

1.2 归一化

单目眼睛图像容易因头部姿态的变化而失真,降低了视线估计的准确性。对此,通过头部姿态坐标系的信息可以建立归一化坐标系以减少眼睛图像失真。

头部姿态坐标系采用EPnP^[16]方法建立。如图3所示(彩色效果见《计算机工程》官网HTML版),从右眼中心到左眼中心的方向是头部姿态坐标系的X轴。与眼睛和嘴巴中心组成的XOY平面垂直的方向是Z轴。白点表示眼角和嘴角的特征点,由白点计算得到的眼睛与嘴巴中心形成的灰色三角形XOY,进而形成头部姿态坐标系。黑色虚线箭头是视线向量 $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$,角度 θ 是视线向量在垂直方向Y上的投影,角度 φ 是视线向量在水平方向X上的投影。



图3 头部姿态坐标系与人眼图像的映射关系

Fig.3 Mapping relationship between head pose coordinate system and eyes photo

图4所示为单目相机坐标系和头部姿态坐标系组成的归一化坐标系。归一化坐标系 S_n 的3个坐标轴为 X_n, Y_n 和 Z_n ,原点为 O_n 。头部姿态坐标系 S_h 的3个坐标轴为 X_h, Y_h 和 Z_h ,原点为 O_h 。相机坐标系 S_c 的3个坐标轴为 X_c, Y_c 和 Z_c ,原点为 O_c 。 O_n, Z_n 与 O_c, Z_c 相同。 X_h^z 是 X_h 到垂直于 Z_n 的平面的投影,如式(6)所示:

$$X_h^z = \frac{X_h \times Z_n}{|X_h| |Z_n|} \times \frac{Z_n}{|Z_n|} \quad (5)$$

X_h^z 是 X_h 到垂直于 Z_n 的平面的投影,如式(6)

所示:

$$X_h^z = X_h - X_h^z \quad (6)$$

Y_h^z 是 Y_h 到 Z_n 的投影,如式(7)所示:

$$Y_h^z = \frac{y_h \times Z_n}{|y_h| |Z_n|} \times \frac{Z_n}{|Z_n|} \quad (7)$$

Y_h^y 是 Y_h 到垂直于 Z_n 的平面的投影,如式(8)

所示:

$$Y_h^y = Y_h - Y_h^z \quad (8)$$

X_n 平行于 X_h^z, Y_n 平行于 Y_h^y 。旋转矩阵 R 是从相机坐标系到归一化坐标系的变换矩阵,可以表示为:

$$R = [X_h^z, Y_h^y, Z_c]^T \quad (9)$$

如图4所示,通过将三维脸投影到归一化坐标系中的焦平面来归一化眼睛图像。

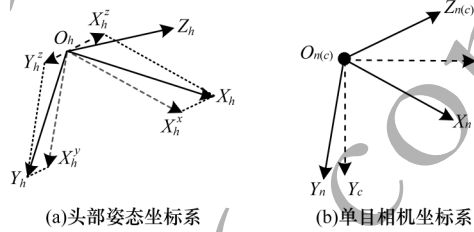


图4 归一化坐标系

Fig.4 Normalization coordinate system

眼睛尺寸也应该通过保持眼球和单目相机之间距离的恒定来归一化。尺度矩阵 S 用于确保从眼睛中心到归一化坐标系原点的距离在 Z_n 上的投影具有相同的值 d_s 。假设归一化坐标系中眼睛中心的坐标为 $(E_{Eycx}, E_{Eycy}, E_{Eycz})$ 。从眼睛中心到归一化坐标系原点的距离在 Z_n 上的投影为 $d = E_{Eycz}$ 。因此,比例矩阵 S 可以表示为:

$$S = D_{\text{diag}}(1, 1, \lambda) = D_{\text{diag}}\left(1, 1, \frac{d_s}{d}\right) = D_{\text{diag}}\left(1, 1, \frac{d_s}{E_{Eycz}}\right) \quad (10)$$

归一化坐标系的构造如图5所示。相机坐标系左乘旋转矩阵 R 以归一化 Z_n 视图的差异,左乘尺度矩阵 S 以统一眼球与单目相机之间的距离。以上2个步骤令相机坐标系构建了归一化坐标系。

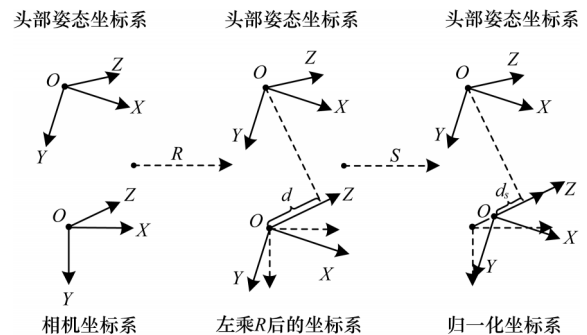


图5 归一化坐标系的构造

Fig.5 Structure of normalization coordinate system

双眼中心的位置是视线估计的基础^[17],采用哈尔(Haar)级联分类器^[18]寻找双眼的中心。为确保眼球位于图像的中心,通过更改相机 C_{in} 内参矩阵中的 C_x 和 C_y 来调整光轴,如式(11)所示:

$$C_{in} = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

最后基于直方图均衡化对图像进行二值化,以减小不同光照条件对图像的影响。

1.3 卷积神经网络的表现提取

研究表明,在进行表观图像提取时,卷积神经网络模型可以产生丰富的表示^[19]。本文选择的卷积神经网络如InceptionV3^[20],在分类任务中具有稳定和高效的性能。在所提方法中,卷积神经网络模型将归一化的眼睛图像映射到垂直和水平方向的视线角度。如图6所示,卷积神经网络模型由卷积层、最大池化层、平均池化层和块组成。卷积神经网络模型的输入是经过4倍放大后的归一化灰度眼部图像,从 36×60 变大到 144×240 。块前的卷积层滤波器尺寸为 3×3 ,块的卷积层滤波器尺寸为 1×1 、 1×3 、 3×1 、 3×3 、 5×5 、 1×7 和 7×1 ,用于提取不同尺度的特征并将它们组合在一起。在每个卷积层后为修正线性单元(Rectified Linear Unit, ReLu)层^[21]和批归一化(Batch Normalization, BN)层^[22]。丢弃(Dropout)层用于避免过度拟合。为减少权值参数,用 1×1 滤波器的卷积层代替全连接层。最后的柔性最大(Softmax)分类层指示视线角度各类别的预测概率。使用交叉熵函数作为损失函数来度量预测视线角度与实际视线角度之间的失配,可表示为:

$$L = -\frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N p_{ij} \log_a q_{ij} \quad (12)$$

其中: M 是样本数; N 是分类数; p_{ij} 是属于分类 j 的样本 i 的实际概率; q_{ij} 是属于分类 j 的样本 i 的预测概率。

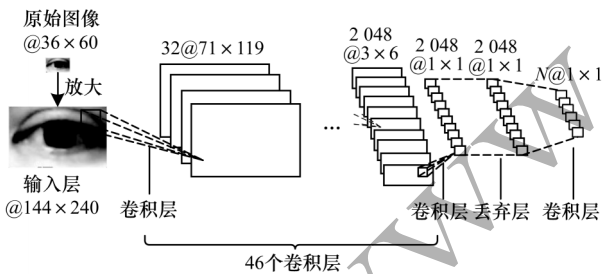


图6 卷积神经网络模型

Fig.6 Convolutional neural network model

1.4 分类

在注视屏幕上的某个点时,视线在垂直方向上的移动并不影响其在水平方向上的移动,因此可以在2个单独任务中估计垂直角 θ 和水平角 φ 。在每个任务中,分类数为 N ,总的视线角度空间为 $[-90^\circ, 90^\circ]$, W 表示类的区间宽度,则分类数 N 可以表示为 $N = \frac{180}{W}$ 。第 i 类中某个眼睛图像的视线角度为 α ,其对应的角度范围为 K_i ,如下:

$$\begin{cases} \alpha \in K_1: \alpha \in \left[-90^\circ, \left(-90 + \frac{W}{2}\right)^\circ\right] | i=1 \\ \alpha \in K_i: \alpha \in \left[\left(-90 + i \times W - \frac{W}{2}\right)^\circ, \left(-90 + i \times W + \frac{W}{2}\right)^\circ\right] \\ | i=2, 3, \dots, N-1 \\ \alpha \in K_N: \alpha \in \left[\left(-90 + N \times W - \frac{W}{2}\right)^\circ, 90^\circ\right] | i=N \end{cases} \quad (13)$$

1.5 误差分析

总的平均角度误差由区间误差和类别误差组成。

设 K_i 类中的视线角 θ_i 在角度空间中均匀分布,则有:

$$\theta_i \sim U\left(-90 + i \times W - \frac{W}{2}, -90 + i \times W + \frac{W}{2}\right) \quad (14)$$

θ_i 对应类别的角度范围 K_i 的中点为 $\hat{\theta}_i$,计算式如式(15)所示:

$$\hat{\theta}_i = -90 + i \times W \quad (15)$$

θ_i 与 $\hat{\theta}_i$ 之间的误差为区间误差,区间误差期望为:

$$m_{\text{interval}}^i = |\theta_i - \hat{\theta}_i| = |\theta_i + 90 - i \times W| = \left| U\left(-\frac{W}{2}, \frac{W}{2}\right) \right| \quad (16)$$

θ_i 的区间误差期望与 i 无关,因此式(16)可以写为:

$$m_{\text{interval}} = \left| U\left(-\frac{W}{2}, \frac{W}{2}\right) \right| = \left| U\left(0, \frac{W}{2}\right) \right| = \frac{W}{4} \quad (17)$$

所有类的区间误差期望与 N 成反比,为:

$$E(m_{\text{interval}}) = \frac{1}{N} \sum_{i=1}^N \frac{W}{4} = \frac{W}{4} \quad (18)$$

实际分类数 N 和预测分类数 \hat{N} 之间的失配为 ΔN :

$$\Delta N = N - \hat{N} \quad (19)$$

由 ΔN 引起的角度误差称为类别误差 $m_{\text{classification}}$,如式(20)所示:

$$m_{\text{classification}} = \Delta N \times L = (N - \hat{N}) \times W = 180 - \hat{N} \times W \quad (20)$$

总的角度误差可以通过区间误差 m_{interval} 和类别误差 $m_{\text{classification}}$ 来计算,如式(21)所示:

$$\begin{cases} m_{\text{total}} = m_{\text{classification}} + m_{\text{interval}} = \\ \left| m_{\text{classification}} \right| + \left| m_{\text{interval}} \right|, m_{\text{classification}} \times m_{\text{interval}} < 0 \\ \left| m_{\text{classification}} \right| - \left| m_{\text{interval}} \right|, m_{\text{classification}} \times m_{\text{interval}} \geq 0 \end{cases} \quad (21)$$

在计算总的角度平均误差时,根据式(21)可知, $|m_{\text{interval}}|$ 可以对消,因此式(21)总的角度平均误差可简化为 $|m_{\text{classification}}|$,即 \hat{N} 的线性函数 f 的绝对值,故可用黄金分割法来寻找最优的分类数 N 。

$$E(m_{\text{total}}) = \overline{m_{\text{total}}} = \overline{|m_{\text{classification}}|} = |f(\hat{N})| \quad (22)$$

1.6 黄金分割法优化

黄金分割法能为本文所提方法寻找测试数据集优化的分类数 N 。分类数 N 的搜索范围设为 $[30, 360]$ 。

算法1 黄金分割

输入 MPIIGaze数据集

输出 最优分类数 N

1. Set the left side and right side of query range as $l_q = 30$ and $r_q = 360$.
2. while $r_q - l_q > 30$ do
3. Set left golden section point as $l_g = (1 - 0.618) \times (r_q - l_q) + l_q$
4. Set $N = l_g$ and use the appearance-based method proposed in this paper to do the classify task to get the mean angular error $\Delta\alpha_l$.
5. Set right golden section point as $r_g = 0.618 \times (r_q - l_q) + l_q$.
6. Set $N = r_g$ and use the appearance-based method proposed in this paper to do the classify task to get the mean angular error $\Delta\alpha_r$.
7. if $\Delta\alpha_l > \Delta\alpha_r$ then
8. $l_q = l_g$
9. else
10. $r_q = r_g$
11. end if
12. end while
13. return l_q

2 实验结果与分析

2.1 数据集的选择

本文选取 MPIIGaze 数据集^[12], 该数据集数据来自具有不同头部姿态的 15 人, 图像分辨率较低, 包括戴或不戴眼镜的男人和女人。为测试归一化测试坐标系的作用, 卷积神经网络模型的输入使用了归一化坐标系形成的 36×60 眼睛图像, 如图 7 所示。经过归一化处理后, 选取最后一个人的 1 464 幅图像作为测试数据, 其余 156 522 幅图像作为训练数据, 以保证模型的泛化。

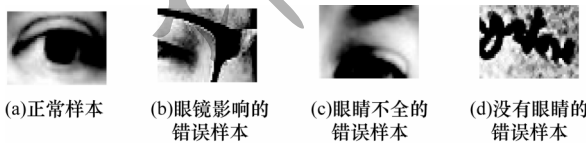


图 7 经过归一化的 MPIIGaze 数据集样本

Fig.7 Normalized sample of MPIIGaze data set

2.2 卷积神经网络模型的训练

使用型号为 GeForce GTX 1080 Ti 的显卡训练卷积神经网络模型时, 最初的学习率设置为 0.01, 共迭代 300 次, 每经过 14 次迭代衰减 6%, 批大小设置为 256, 采用 ADAM^[23] 优化器。实验结果表明优化后平均每幅图像处理耗时约为 0.18 s。

2.3 评价

2.3.1 对区间误差的评价

如 3.6 节所述, 区间误差期望与分类数成反比。图 8 所示为不同分类数 N 在测试数据集上的区间误差期望。带 \star 虚线表示垂直方向视线估计的区间误差期望, 带 \times 虚线表示水平方向视线估计的区间误差期望。图 9 所示为在总角度误差中的区间误差分量期望, 对于 MPIIGaze 数据集已训练的卷积神经网络模型, 按 1~256 不同测试样本数量进行测试, 可以看到区间误差分量期望的影响较小。

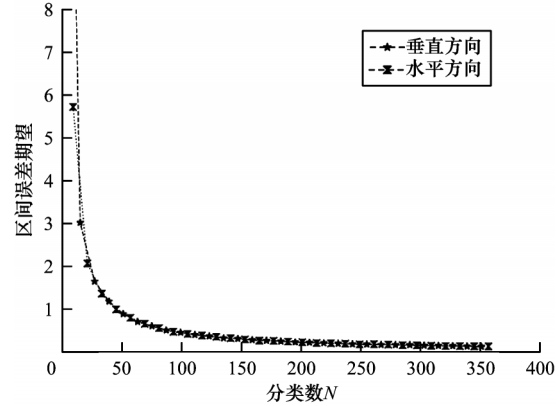


图 8 区间误差期望变化

Fig.8 Expected change in interval error

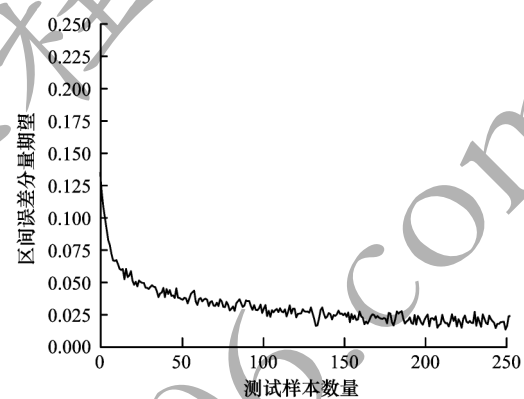


图 9 区间误差分量期望变化

Fig.9 Expected change in interval error component

2.3.2 对误差的评价

误差受分类数 N 的影响。使用黄金分割法寻找优化分类数 $N=180$, 迭代 4 次:

$$\begin{aligned}
 (l_q, l_g, r_g, r_q) &= (30, 150, 240, 360) | \Delta\alpha_{150} < \Delta\alpha_{240} \\
 (l_q, l_g, r_g, r_q) &= (30, 90, 150, 240) | \Delta\alpha_{150} < \Delta\alpha_{90} \\
 (l_q, l_g, r_g, r_q) &= (90, 150, 180, 240) | \Delta\alpha_{180} < \Delta\alpha_{150} \\
 (l_q, l_g, r_g, r_q) &= (150, 180, 210, 240) | \Delta\alpha_{180} < \Delta\alpha_{210} \quad (23)
 \end{aligned}$$

其中: (l_q, l_g, r_g, r_q) 表示在每次迭代中黄金分割法的中间变量。

表 1 所示为双眼在垂直和水平方向上的预测角度和实际角度之间的平均角度误差。

表 1 双眼在垂直和水平方向上的平均角度误差

Table 1 Average angle error of both eyes in vertical and horizontal direction ($^{\circ}$)

指标	$N=30$	$N=90$	$N=150$	$N=180$	$N=210$	$N=240$	$N=360$
$\Delta\theta_l$	3.455	3.406	3.454	3.303	3.379	3.534	3.873
$\Delta\phi_l$	2.600	2.615	2.529	2.506	2.643	2.835	2.859
$\Delta\theta_r$	4.308	4.293	3.771	3.669	3.726	3.844	3.800
$\Delta\phi_r$	3.377	3.095	3.199	3.057	3.152	3.184	3.324

按式(4)合成 $\Delta\theta$ 、 $\Delta\phi$, 得到预测的视线向量和实际的视线向量间的平均角度误差 $\Delta\alpha$, 如表 2 所示。

表 2 双眼的平均角度误差

Table 2 Average angle error of binocular (°)							
指标	N=30	N=90	N=150	N=180	N=210	N=240	N=360
$\Delta\alpha_l$	4.323	4.293	4.280	4.145	4.289	4.529	4.813
$\Delta\alpha_r$	5.471	5.139	4.944	4.744	4.879	4.990	5.047

图 10 所示为黄金分割法搜索的分类数 N 的情况,其中,虚线表示平均角度误差为 4.8° 。分类数 N 的平均角度误差形状是凹的,当 N 设置为 180 时,本文所提方法获得了左眼 4.145° 、右眼 4.744° 、平均 4.445° 的优化性能。

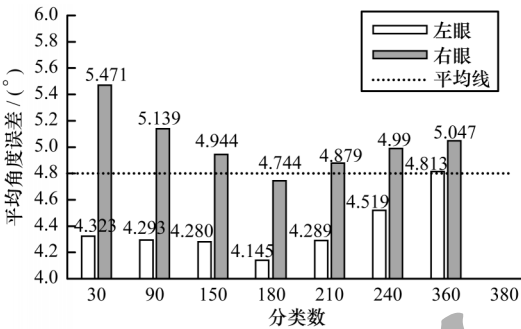


图 10 黄金分割法搜索优化的分类数

Fig.10 The number of classification optimized by the search of the golden section method

2.3.3 对归一化坐标系的评价

用文献[8]中所提归一化方法和本文所提归一化坐标系所输出的不同眼睛图像作为同一卷积神经网络模型的输入,评价归一化坐标系方法。当分类数 N 设置为 180 时,比较结果如表 3 所示。

表 3 不同归一化方法在垂直和水平方向的平均角度误差

Table 3 Average angle errors of different normalization methods in vertical and horizontal directions (°)		
指标	文献[12]方法	本文方法
$\Delta\theta_l$	3.450	3.303
$\Delta\varphi_l$	2.512	2.506
$\Delta\theta_r$	3.752	3.669
$\Delta\varphi_r$	3.203	3.057

与文献[8]所提方法相比,本文建立的归一化坐标系方法,左眼和右眼的误差分别从 4.267° 和 4.932° 降低到 4.145° 和 4.744° 。

2.3.4 整体评价

表 4 是本文所提方法与传统基于表观方法的对比。本文所提方法结果为左眼 4.145° 、右眼 4.744° ,表现最优。与文献[13-15]基于表观的方法相比,该方法平均角度误差从 4.8° 降低到 4.445° ,误差降低了 7.4%。

表 4 不同基于表观方法的平均角度误差比较

Table 4 Comparison of average angle error of different appearance-based methods (°)					
指标	文献[12]方法	文献[13]方法	文献[14]方法	文献[15]方法	本文方法
$\Delta\alpha$	6.3	5.4	5.4	4.8	4.445

3 结束语

本文提出一种基于表观的归一化坐标系分类视线估计方法,通过校准相机坐标系使眼球与单目相机的距离保持恒定,从而增强眼睛图像处理的效果及减小不同头部姿态对图像的影响。同时,通过构建卷积神经网络模型,并使用黄金分割法进行最终优化,提升视线检测的准确率。实验结果表明,相比已公开的同类方法,在 MPIIGaze 数据集中,本文所提方法使视线估计误差平均约降低了 7.4%。由于眼睛图像容易受到光照、头部姿态等综合因素的影响,下一步将围绕人脸光照补偿展开研究,提高视线检测效果。

参考文献

[1] POWELL W A, CORBETT N, POWELL V. Computational linguistics: concepts, methodologies, tools, and applications [J]. Express Shipping, 2020, 34(2): 1459-1488.

[2] SUFI F B, GAZZANO J D D, CALLE F R, et al. Multi-camera tracking system applications based on reconfigurable devices: a review [C]//Proceedings of 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering. Washington D. C., USA: IEEE Press, 2019: 1-5.

[3] PARK S H, YOON H S, PARK K R. Faster R-CNN and geometric transformation-based detection of driver's eyes using multiple near-infrared camera sensors [J]. Sensors, 2019, 19(1): 197-205.

[4] SELIM M, FIRINTEPE A, PAGANI A, et al. Autopose: large-scale automotive driver head pose and gaze dataset with deep head orientation baseline [C]//Proceedings of the 15th International Conference on Computer Vision Theory and Applications. Washington D. C., USA: IEEE Press, 2020: 599-606.

[5] LEE G J, JANG S W, KIM G Y. Pupil detection and gaze tracking using a deformable template [J]. Multimedia Tools and Applications, 2020, 9(11): 13-20.

[6] MAHMUD S, LIN X, KIM J H. Interface for human machine interaction for assistant devices: a review [C]//Proceedings of the 10th Annual Computing and Communication Workshop and Conference. Washington D. C., USA: IEEE Press, 2020: 768-773.

[7] WANG Z, ZHAO J, LU C, et al. Learning to detect head movement in unconstrained remote gaze estimation in the wild [C]//Proceedings of IEEE Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2020: 3443-3452.

[8] SUGANO Y, MATSUSHITA Y, SATO Y. Learning-by-synthesis for appearance-based 3D gaze estimation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1821-1828.

[9] BREIMAN L. Random forests, machine learning 45 [J]. Journal of Clinical Microbiology 2001, 2(30): 199-228.

- [10] COVER T. Estimation by the nearest neighbor rule[J]. IEEE Transactions on Information Theory, 1968, 14(1): 50-55.
- [11] LU F, SUGANO Y, OKABE T, et al. Inferring human gaze from appearance via adaptive linear regression [C]// Proceedings of International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2011: 153-160.
- [12] ZHANG X, SUGANO Y, FRITZ M, et al. Appearance-based gaze estimation in the wild[C]//Proceedings of IEEE Conference on Computer Vision and Pattern recognition. Washington D. C. , USA: IEEE Press, 2015: 4511-4520.
- [13] ZHANG X, SUGANO Y, FRITZ M, et al. Mpiigaze: real-world dataset and deep appearance-based gaze estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 41(1): 162-175.
- [14] RANJAN R, DE MELLO S, KAUTZ J. Light-weight head pose invariant gaze tracking [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C. , USA: IEEE Press, 2018: 2156-2164.
- [15] ZHANG X, SUGANO Y, FRITZ M, et al. It's written all over your face: full-face appearance-based gaze estimation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 51-60.
- [16] LI M, CHEN R, LIAO X, et al. A precise indoor visual positioning approach using a built image feature database and single user image from smartphone cameras [J]. Remote Sensing, 2020, 12(5): 869-873.
- [17] YIFAN X, JIANWEN L, JUNYU D, et al. Hybrid regression and isophote curvature for accurate eye center localization[J]. Multimedia Tools and Applications, 2020, 79(2): 805-824.
- [18] SINGH A, HERUNDE H, FURTADO F. Modified haar-cascade model for face detection issues[J]. International Journal of Research in Industrial Engineering, 2020, 38(12): 145-1677.
- [19] DHILLON A, VERMA G K. Convolutional neural network: a review of models, methodologies and applications to object detection [J]. Progress in Artificial Intelligence, 2020, 9(2): 85-112.
- [20] BENDJILLALI R I, BELADGHAM M, MERIT K, et al. Illumination-robust face recognition based on deep convolutional neural networks architectures [J]. Indonesian Journal of Electrical Engineering and Computer Science, 2020, 18(2): 1015-1027.
- [21] NANNI L, LUMINI A, GHIDONI S, et al. Stochastic selection of activation layers for convolutional neural networks [J]. Sensors, 2020, 20(6): 1626-1634.
- [22] WANG S H, MUHAMMAD K, HONG J, et al. Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization [J]. Neural Computing and Applications, 2020, 32(3): 665-680.
- [23] LUO J, LIU J, LIN J, et al. A lightweight face detector by integrating the convolutional neural network with the image pyramid [J]. Pattern Recognition Letters, 2020, 133(5): 180-187.

编辑 赖玉玲