

头部姿势估计研究综述*

唐云祁^{1,2} 孙哲南¹ 谭铁牛¹

¹(中国科学院自动化研究所 北京 100190)

²(中国人民公安大学 刑事科学技术学院 北京 100038)

摘 要 头部姿势估计的研究成果众多,但缺乏系统、详尽的综述性文献.针对该问题,首先介绍头部姿势的表达方式和头部姿势估计的基本问题,然后从方法依赖的数据源、人工干预程度和实现原理等角度对已有头部姿势估计方法进行分类,并以实现原理为主线,以所依赖的数据源为辅助,对头部姿势估计问题的研究文献进行详细整理和分析.

关键词 头部姿势估计,模式识别,计算机视觉,人脸识别

中图法分类号 TP 391.4

A Survey on Head Pose Estimation

TANG Yun-Qi^{1,2}, SUN Zhe-Nan¹, TAN Tie-Niu¹

¹(*Institute of Automation, Chinese Academy of Sciences, Beijing 100190*)

²(*Department of Criminal Science and Technology, People's Public Security University of China, Beijing 100038*)

ABSTRACT

Lots of head pose estimation methods have been presented, while there is not any survey that systematically summarizes the existing methods. To address this problem, detailed description of head pose estimation is firstly presented to define what head pose estimation is and how to describe it. Then, the existing head pose estimation methods are categorized according to the data source used by the methods, the automation of the methods and the principle of the methods. Furthermore, the principle and data source are taken as the main line and auxiliary line respectively to detailedly review and analyze the existing head pose estimation methods.

Key Words Head Pose Estimation, Pattern Recognition, Computer Vision, Face Recognition

* 中国科学院战略性先导科技专项项目(No. XDA06030300)、国家“973”重点基础研究发展计划项目(No. 2012CB316300) 资助

收稿日期: 2013-04-30; 修回日期: 2013-06-15

作者简介 唐云祁(通讯作者)男,1983年生,博士,讲师,主要研究方向为生物特征识别、模式识别. E-mail: yunqi.tang@gmail.com. 孙哲南男,1976年生,博士,副研究员,主要研究方向为生物特征识别、模式识别、计算机视觉. 谭铁牛男,1965年生,研究员,博士生导师,主要研究方向为生物特征识别、模式识别、计算机视觉.

1 引言

在计算机视觉领域,头部姿势估计(Head Pose Estimation)指的是计算机以用户所成图像来估计其头部姿态参数的过程.

近年来,头部姿势估计成为模式识别和计算机视觉领域内的活跃课题.一方面是因为头部姿势估计在人脸识别、视线估计^[1]等头部相关研究中有着广泛的应用前景.当前计算机视觉研究工作的基本方法是利用目标所成图像的特性(包括颜色、形状、纹理等)来解决现实应用所面临的问题,如目标检测、识别等.然而,目标物体在摄像机中所呈现的图像受姿态的影响较大,同一物体在不同姿态下所呈现图像的颜色、形状、纹理等信息都不尽相同.头部是人脸、虹膜、眼睛的载体,当头部姿势发生改变时,这3种器官在摄像机中所成图像包含的信息也必然会发生改变,这势必影响人脸识别、虹膜识别、视线估计和表情识别方法的研究.因此,研究头部姿势估计对人脸识别、虹膜识别、视线估计和表情识别的研究工作有着重要的现实意义.

另一方面是因为头部姿势估计具有重要的学术研究价值.头部姿势估计是让计算机从采集到的视频图像中估计出用户头部的姿态参数,这是一个典型而又复杂的模式识别问题.解决此类问题的挑战性在于:1) 头部姿势估计涉及头部图像分割问题.图像分割一直都是数字图像处理领域的一个难题,目前还没有统一的解决方案.受光照变化、复杂背景和姿态变化等因素的影响,头部图像的精确分割是一个极具挑战的难题.2) 头部姿势估计涉及特征表达设计问题.图像特征表达设计是模式识别领域的一个开放性话题,目前研究者在“怎样的特征是图像的本质表达”问题上还未达成共识.设计一种统一的、能精确描述头部姿势变化规律的特征表达方法也极具挑战.

目前,国内外关于头部姿势估计问题的研究众多,但这方面的综述文章却较少. Murphy-Chutorian 等^[2]2009年发表关于头部姿势估计的综述性论文.这是头部姿势估计研究领域的第一篇、也是唯一的一篇综述性文献.目前来看,该文献至少存在如下3个方面的局限性:1) 没有给出头部姿势估计问题的详细描述;2) 综述的文献大多是基于RGB彩色图像的头部姿势估计方法,缺乏对基于深度图像或者三维图像头部姿势估计方法的总结和比较;3) 分类方法有一定的局限性.本文将在详细描述头部姿势估计问题的基础上,对头部姿势估计领域的主流方法进行系统总结.

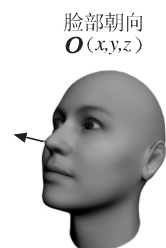
2 头部姿势估计问题描述

2.1 头部姿势表达方式

一般来说,头部姿势有两种表达方式:脸部朝向(Face Orientation)和欧拉旋转角(Euler Rotation Angles).

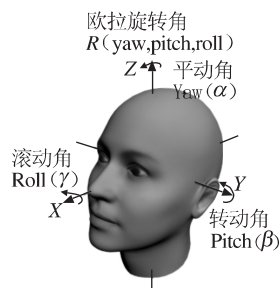
2.1.1 头部姿势的脸部朝向表达方式

脸部朝向就是脸部所朝向的方向,可用三维空间中某一坐标系下的方向向量 $O(x, y, z)$ 表示,如图1(a)所示.为便于理解,可选择如图1(c)中所示 xyz 坐标系来作为脸部朝向的参考坐标系.那么,图1(c)所示正脸图像的脸部所朝向则为 $O(1, 0, 0)$.



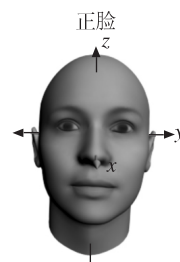
(a) 脸部朝向表达方式

(a) Face orientation model



(b) 欧拉旋转角表达方式

(b) Euler rotation angle model



(c) 正脸图像

(c) Frontal face image

图1 头部姿势的两种表达方式

Fig.1 Two description models of head pose

脸部朝向是对头部姿态最直接和自然的表达, 在一些应用背景下(如视线估计等), 头部姿势估计的目的就是要估计出脸部朝向. 但脸部朝向并不是头部姿势的完整表达, 不能唯一表达头部姿态的所有状态. 如图 2(a) 为正脸图像, (b) 和 (c) 是正脸图像以 X 轴为旋转轴分别顺时针和逆时针旋转一定角度后得到的姿态图像. 从图中可知, 3 幅图像的脸部朝向都是 $O(1, 0, 0)$, 但这 3 幅图像的头部姿势显然各不相同.

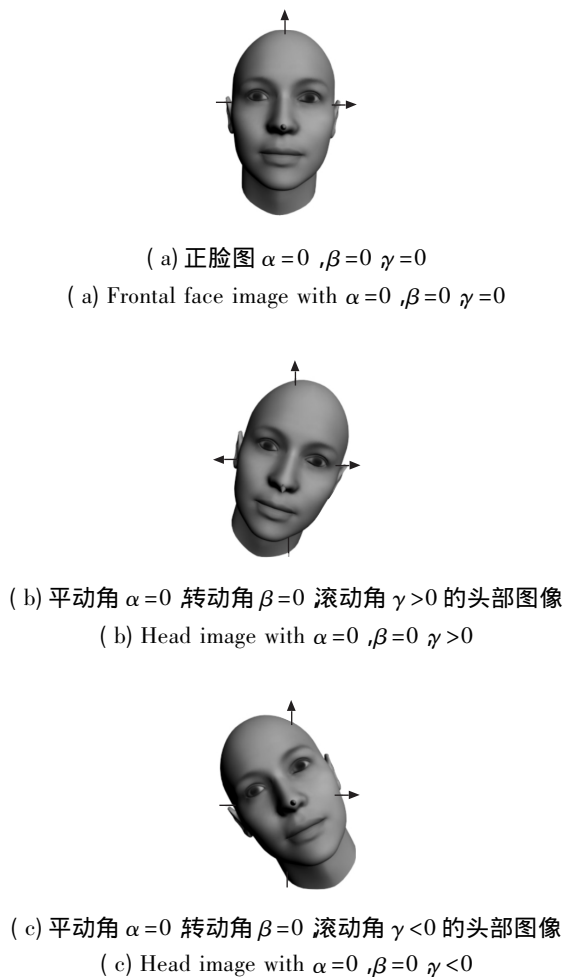


图 2 相同脸部朝向时不同头部姿势示意图

Fig. 2 Sketch map of different head poses with same face orientation

2.1.2 头部姿势的欧拉旋转角表达方式

欧拉旋转角能精确表达所有状态下的头部姿态. 欧拉角是用来描述刚体在三维欧几里得空间取向的一组三维角参量, 由平动角(Yaw)、转动角(Pitch)、滚动角(Roll)组成, 为欧拉首先提出而得名.

欧拉角一般涉及两个坐标系: 参考坐标系和刚体坐标系. 参考坐标系可以是三维空间里的任意坐标系, 其本身是静止不动的, 不随刚体的旋转而改变

位置. 而刚体坐标系则固定于刚体上, 随着刚体的旋转而旋转. 为便于理解, 取如图 1(b) 所示 XYZ 坐标系作为头部姿势欧拉旋转角表达的刚体坐标系, 如图 1(c) 所示脸部朝向参考坐标系 xyz 作为头部姿势欧拉旋转角表达的参考坐标系.

当头部图像为正脸时(如图 1(c) 所示), 刚体坐标系和脸部朝向参考坐标系重叠, 此时头部的欧拉旋转角为 $R(0, 0, 0)$, 脸部朝向为 $O(1, 0, 0)$. 那么对于头部的任意一个姿态 $R(\alpha, \beta, \gamma)$, 如图 1(b) 所示, 可由正脸头部依次围绕刚体坐标轴 ZXZ 旋转 α, β, γ 角值而得到.

2.1.3 脸部朝向和欧拉旋转角的关系

脸部朝向和欧拉旋转角是头部姿势的两种不同表达, 都由 3 个参量组成. 但脸部朝向所包含的信息量较欧拉旋转角少一个维度, 仅能表达欧拉旋转角中的平动角(Yaw)和转动角(Pitch). 这是因为, 刚体坐标 X 轴方向即为脸部朝向, 如图 1(b) 所示, 无法指示头部绕 X 轴旋转的滚动角(Roll)信息. 因此, 在已知脸部朝向的情况下, 仅得到欧拉旋转角的有效平动角和转动角分量.

下面是根据脸部朝向计算欧拉旋转角的平动角和转动角分量的计算过程.

根据脸部朝向计算欧拉旋转角的平动角和转动角分量需要借助四元数的辅助. 四元数(Quaternion)是 Hamilton 于 1843 年提出的数学概念, 可表示矢量和物体的旋转. 四元数的基本定义^[3]如下: 设 Q 是实数域上的四维向量空间, 正交基底 $(1, i, j, k)$, $(0, 1, i, j)$, $(0, i, 1, j)$, $(0, j, i, 1)$ 分别用 e, i, j, k 表示, 则 Q 中的元素均可表示为

$$q = xi + yj + zk + we,$$

其中 q 为一个四元数.

四元数 $Q(x, y, z, w)$ 的物理意义表示, 方向向量绕给定轴的一个旋转. 若旋转轴为 $A(a_x, a_y, a_z)$, 旋转角度为 θ , 则四元数中的 x, y, z, w 分别为

$$\begin{aligned} x &= \sin(\theta/2) \cdot a_x, \\ y &= \sin(\theta/2) \cdot a_y, \\ z &= \sin(\theta/2) \cdot a_z, \\ w &= \cos(\theta/2). \end{aligned}$$

四元数和欧拉旋转角之间存在着一一映射关系. 给定四元数 $Q(x, y, z, w)$, 其相应欧拉旋转角 $R(\alpha, \beta, \gamma)$, 计算如下:

$$\begin{bmatrix} \gamma \\ \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \arctan \frac{2(wx + yz)}{1 - 2(x^2 + y^2)} \\ \arcsin(2(wy - zx)) \\ \arctan \frac{2(wz + xy)}{1 - 2(y^2 + z^2)} \end{bmatrix}.$$

根据上述四元数到欧拉旋转角的转换关系,可依据给定脸部朝向 $B(b_x, b_y, b_z)$, 计算其欧拉旋转角. 如图 3 所示, 脸部朝向 $B(b_x, b_y, b_z)$ 可视为方向向量 $(1, 0, 0)$ 绕 $OA(a_x, a_y, a_z)$ 旋转 θ 后所得. 其中 OA 为方向向量 $(1, 0, 0)$ 和脸部朝向 $B(b_x, b_y, b_z)$ 的法向量,

$$a_x = 0, \quad a_y = \frac{b_z}{\sqrt{b_y^2 + b_z^2}}, \quad a_z = \frac{-b_y}{\sqrt{b_y^2 + b_z^2}},$$

θ 为方向向量 $(1, 0, 0)$ 和脸部朝向 $B(b_x, b_y, b_z)$ 的夹角 $\theta = \arccos(b_x)$. 代入上式便可得到脸部朝向对应的欧拉旋转角分量.

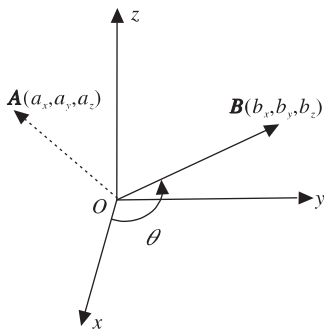


图 3 根据脸部朝向计算欧拉旋转角分量示意图

Fig. 3 Calculation of Euler rotation angles according to face orientation

通过上式,可实现从脸部朝向到欧拉旋转角平动角和转动角分量的转换. 同样,在已知欧拉旋转角 $R(\alpha, \beta, \gamma)$ 的情况下,也可得到其相应的脸部朝向.

依据欧拉旋转角计算脸部朝向的实质就是计算刚体坐标系下点 $P(1, 0, 0)$ 在脸部朝向参考坐标系下的坐标 $O(x, y, z)$. 根据欧拉角的动态定义,最初,脸部朝向参考坐标轴 xyz 与刚体坐标轴 XYZ 重叠,依次让头部绕着 z 轴旋转 γ 角值,绕着 x 轴旋转 β 角值,绕着 z 轴旋转 α 角值,便可得脸部朝向 $O(x, y, z)$:

$$O = z(\alpha) \cdot x(\beta) \cdot z(\gamma) \cdot P, \quad (1)$$

其中 $z(\alpha)$ 表示头部绕参考坐标系 z 轴旋转 α 角值, $z(\gamma)$ 表示头部绕参考坐标系 z 轴旋转 γ 角值, $x(\beta)$ 表示头部绕参考坐标 x 轴旋转 β 角值. 用旋转矩阵表示:

$$z(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$x(\beta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\beta) & -\sin(\beta) \\ 0 & \sin(\beta) & \cos(\beta) \end{bmatrix},$$

$$z(\gamma) = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

将上述 3 个公式代入式(1)便可得到脸部朝向 $O(x, y, z)$.

2.2 头部姿势估计所要解决的问题

如 2.1 节所述,头部姿势可用一个三维向量来描述. 那么头部姿势估计要解决的问题就是从图像中估计出用户的头部姿态,从而得到一个三维姿态参数. 三维姿态参数可以是脸部朝向,也可以是欧拉旋转角.

具体来说,头部姿势估计要解决的问题是寻找从图像空间到三维姿态空间(或从图像空间到特征空间再到三维姿态空间)的映射关系,如图 4 所示.

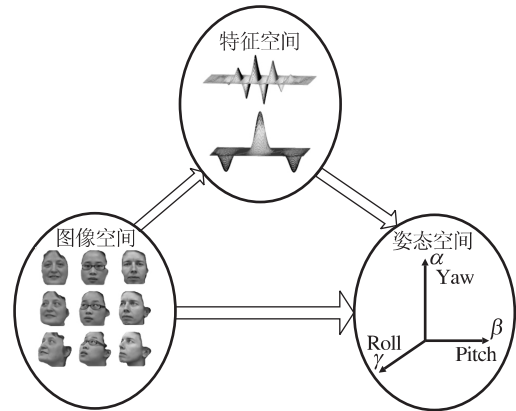


图 4 头部姿势估计问题描述示意图

Fig. 4 Graphic description of head pose estimation problem

3 头部姿势估计方法分类

头部姿势估计问题的复杂性在某种程度上导致头部姿势估计方法的多样性. 20 世纪 90 年代以来,国际上提出了几十余种头部姿势估计方法,形成多种方法并存的局面. 下面将对当前国内外的主流头部姿势估计方法进行分类综述.

表 1 为现有头部姿势估计方法的总体分类情况. 根据所依赖数据源的不同,大致可将这些方法分为基于二维彩色图像的方法^[4-8]、基于深度图像的方法^[9-11, 12-18]和基于三维图像的方法^[19]三大类. 第一类方法,也可说是较为传统,使用最为普遍的方法——基于二维彩色图像的头部姿势估计方法. 这类方法不依赖任何特殊硬件,只需一枚普通网络摄

像头便可正常工作,因此易于推广使用。相关研究成果产出最多。但是,这类方法受光照变化、表情变化和遮挡的影响较大。相较而言,第二类方法,基于深度图像的头部姿势估计方法,能较好地克服光照变化和表情变化带来的影响,是头部姿势估计研究的理想之选。特别是近年来廉价深度信息传感器 Kinect 的出现,使得基于深度图像的头部姿势估计研究方法受到研究者们的青睐。第三类方法,基于三维图像的头部姿势估计方法,这方面的相关研究甚少。其原因一方面是由于三维图像传感器较为昂贵,应用面太窄;另一方面是由于从普通摄像头所获取的二维彩色图像中构建三维图像的技术尚不成熟。但三维图像包含任何视角下场景的彩色纹理信息和深度信息,可用数据量较大,使用三维图像来解决计算机视觉问题是将来的发展趋势,随着三维图像采集设备廉价化和三维重建技术的成熟化,基于三维图像的方法必将成为头部姿势估计的新潮流。

根据人工干预程度的不同,还可分为全自动头

部姿势估计方法和半自动头部姿势估计方法^[20-24]。全自动头部姿势估计方法不需人工操作干预,系统能自动给出用户头部姿态参数。当前大多头部姿势估计方法都为全自动头部姿势估计方法。半自动头部姿势估计方法需人工手动处理(如手动标定脸部关键点)之后,系统才能计算出最终姿态参数。

根据实现原理和方式的不同,头部姿势估计方法又可分为基于形状模板的方法^[4-11]、基于检测的方法^[20-25]、基于脸部关键点几何关系的方法^[21-24,26-32]、基于特征回归的方法^[12-18,33-45]、基于子空间学习的方法^[46-54]和基于局部约束模型的方法^[55-56]等 6 大类。

事实上,一个成功的头部姿势估计方法,往往综合考虑各种因素,难以从单个角度对所有方法进行绝对划分。为了能够细致地介绍分析现有头部姿势估计方法,下面本文以方法的实现原理和方式为主线,以方法所依赖的数据源为辅助,对现有头部姿势估计方法进行分类综述。

表 1 头部姿势估计分类情况
Table 1 Categories of head pose methods

分类依据	方法类别	文献	优点	缺点	备注
依赖数据源的不同	基于二维彩色图像的方法	[4]~[8], [20]~[24], [26]~[32], [33]~[45]等	不依赖特殊硬件;易推广	受光照变化、表情变化和遮挡的影响较大	传统主流方法
	基于深度图像的方法	[9]~[11], [12]~[18]等	对光照变化和表情变化较为鲁棒;估计精度高;实时性好	缺失彩色信息	热点方法
	基于三维图像的方法	[19]等	包含各个视角下彩色纹理信息和深度信息,可用信息量大	传感器较为昂贵;三维重建技术尚不成熟;应用面较窄	将来的发展趋势
人工干预程度的不同	全自动头部姿势估计方法	绝大部分方法	符合发展潮流	—	
	半自动头部姿势估计方法	[21]~[24]等	—	已成为历史	
实现原理和方式的不同	基于形状模板的方法	[4]~[11]等	不需要训练学习	较大误差;时间复杂度高、效率低;仅能得到有限离散姿态值	传统方法
	基于检测的方法	[20]~[25]等	不需要头部检测和定位的预处理步骤	姿态检测子的训练工作过于繁琐;系统实时性很难得到保证	传统方法
	基于脸部关键点几何关系的方法	[21]~[24], [26]~[32]等	方法简单,时间复杂度低	脸部关键点的检测和定位是其瓶颈所在	传统方法
	基于特征回归的方法	[12]~[18], [33]~[45]等	实时性好、精度高	受头部检测定位结果的影响较大	主流方法
	基于子空间学习的方法	[46]~[54]等	方法新颖	时间复杂度高,精度比较低	新兴方法
	基于局部约束模型的方法	[55]、[56]等	对光照、形变和遮挡的鲁棒性较强	时间复杂度高;仅能得到有限离散姿态值	新兴方法

3.1 基于形状模板的方法

模板匹配是最早出现的、最为基本的模式识别方法,其数学模型较简单,只需对每个类别建立一个或多个标准模板,分类决策时将待识别的样本与已建立的标准模板进行比对,根据与模板的匹配程度,将样本划分到最相似的类别中。由于这种方法具有直接、简单、在类间差距大时仍可使用的特性,被广泛应用于模式识别领域的各个方向,如目标检测、字符识别、车牌识别等。

1994年 Deymer^[4]把模板匹配的思想引入头部姿势估计。首先把头部姿势空间量化为若干离散点,并为每个姿态准备若干模板图像;然后将待估计头部图像与模板库中的图像一一比较,匹配分数最高的模板图像所对应的头部姿态,即为测试图像的估计结果。

此类方法具有两大优点:1) 可根据现实需要,动态调整模板数量。一般来说,模板数量越多,系统精度越高,但同时耗费的计算资源也就越多;2) 不需训练学习,无需准备训练用的正负样本。与基于学习的方法相比,这样既能避免繁杂的训练样本标定工作,又能避免训练样本选取不当造成的过拟合、泛化性能差的问题。

当然,除上述两大优点,基于形状模板的头部姿势估计方法也同样存在缺陷。

1) 此类方法的头部姿势表达能力有限。因为头部姿态空间是连续的三维空间,而此类方法仅能得到有限离散姿态值,不能完全表达头部在所有状态下的姿势。在现实应用中,研究者往往通过增加模板数量,提高离散姿态点的密度来加强对整个姿态空间覆盖率。这种做法虽然在一定程度上加强了系统头部姿态表达能力,但也增加了模板数量,这样会直接降低系统效率。

2) 此类方法的系统时间复杂度高、效率低。从系统精度方面考虑,模板数量越大,系统精度越高。然而,系统时间复杂度与模板数量成正比,随着模板数量的不断增加,系统时间复杂度会大大提高,随之也会大幅降低系统效率。为克服此弱点,Ng等^[5-6]通过训练一系列支持向量机(Support Vector Machine, SVM)来检测和定位各种姿态下的人脸,并使用所得支撑向量作为模板来估计头部姿势。

3) 此类方法易造成较大误差。此类方法的前提是假设两幅相似头部图像的姿态也是相似的,然而现实情况下这个假设往往是不成立的。如来自同一人的两幅头部图像,如果它们的头部姿态差别不大,那么它们的相似匹配度,极有可能要大于来自不同

两个人的同一姿态下的两幅头部图像。另外,一幅左侧脸和一幅右侧脸的相似程度也极可能大于一幅侧脸和一幅正脸的相似度等。出现类似的情况,就易造成较大误差。为降低系统误差,很多方法试图通过预处理来为此前提假设提供成立的条件。如通过拉普拉斯-高斯滤波器就可保留并加强头部的轮廓信息,去除一些纹理噪声信息,消除不同人所带来的影响。同样,也可使用 Gabor 小波滤波器来提取脸部的方向特征^[7-8],如鼻子的垂直线条信息、嘴的水平线条信息等。

近年来,GPU技术和深度成像技术的发展和普及,为基于形状模板的头部姿势估计方法开辟了新的途径。Breitenstein等^[9]利用GPU的计算性能提出第一个基于深度图像的实时头部姿势估计系统。首先,离线准备各姿态下的脸部模板;然后,利用深度图像的几何特性,检测新来测试图像的鼻尖候选点集;接着,以鼻尖为对准点,把待测试图像与模板图像对齐,并计算其相似度;最终,与待测试图像相似度最高的模板所对应的头部姿势即为所得结果。Breitenstein等的方法为头部姿势估计开启了实时新篇章,此后关于头部姿势的研究大多是以实时系统为目标。在 Breitenstein 方法的基础上,Padeleris等^[10]采用粒子群优化的方法设计一个精度更高的头部姿势估计系统。但此系统需在线手动标定标准参照帧,是一种半自动的头部姿势估计方法。

3.2 基于检测的方法

人脸检测一直是模式识别领域的热门方向,关于它的研究成果数量较多,特别是 Viola等^[57]提出的基于 Boosting 算法,把正脸人脸检测推向高实时、高准确率阶段。

头部姿势估计可看成一个多姿态人脸检测问题。人脸检测的基本思想和方法可以方便地扩展到头部姿势估计上。早在1998年 Huang等^[25]就用人脸检测的思想来做头部姿势估计。此类方法的基本原理如下:首先在离线状态下训练一系列各种姿态下的头部检测子;在线测试时,各头部检测子按序依次扫描待测试图像,响应最大检测子所对应的头部姿势即为估计结果。基于检测的头部姿势估计方法的思想与基于形状模板方法有些类似,都是通过评价待测试图像与参照物的匹配或响应程度,选取最为匹配或响应最大参照物的姿态来作为估计结果。不同的是基于形状模板的头部姿势估计方法通过直接比较两幅图像来实现,而基于检测的头部姿势估计方法则是通过姿态检测子的检测来实现。

基于检测的头部姿势估计方法有两大优点。

1) 不需要头部检测和定位的预处理步骤. 预处理是模式识别方法的重要辅助步骤, 往往又是系统瓶颈所在. 此类方法不需要单独预处理步骤, 集头部检测与头部姿势估计于一体, 回避了预处理所要面对的问题. 2) 与基于形状模板的头部姿势估计方法相比, 此类方法采用机器学习的基本原理学习得到姿态检测子, 更加鲁棒.

当然, 此类方法也具有一些弱点.

1) 姿态检测子的训练工作过于繁琐. 每个姿态检测子的训练都需收集大量的正负训练样本, 这不仅是一个繁杂的工作, 而且随着姿态数量的增加往往易产生一些系统问题. 如在收集两个相近姿态的训练样本时, 若一幅头部图像是一个姿态的正样本, 那么它必定是另外一个姿态的负样本. 然而当前检测算法尚不能从差别不大的正负样本中学习得到一个较好的模型. 在现实应用系统中, 一个维度上的姿态间隔值一般在 30° 以上, 也就是一个维度上姿态检测子不超过 12 个, 这就导致这类方法的估计精度较低.

2) 此类方法的时间复杂度随着检测子数量的增加而线性增长, 系统实时性较难得到保证. 为均衡时间复杂度和检测子数量之间的关系, Rowley 等^[20]提出使用检测子路由的二级检测策略. 对待检测窗口, 首先使用路由由分类器检测到脸部所在位置, 并初步给出其姿态候选结果; 然后, 使用候选姿态相应的检测子进行最终验证, 得到检测结果.

3.3 基于脸部关键点几何关系的方法

借鉴人的感知和思维机理是模式识别和人工智能研究方法中的一个重要分支. 同样, 借鉴人类视觉系统对头部姿势的感知机理也是头部姿势估计的一类重要方法.

研究表明人类视觉系统通过脸部的细节信息(如鼻子的偏转度、脸部的对称度、脸部轮廓在头部轮廓中的相对位置等)来感知头部姿态情况^[21].

基于脸部关键点几何关系的头部姿势估计方法借鉴人类视觉对头部姿势的感知机理, 利用脸部关键点的相对位置来估计头部姿势. 这类方法的基本流程为首先确定脸部关键点所在位置, 然后通过关键点的相对位置来估计头部姿势.

精准定位脸部关键点是此类方法的第一个关键步骤, 也是难点所在. 受光照变化、姿态变化、表情变化及遮挡等诸多因素的影响, 从一张脸部图像上精确定位关键点, 尤其是从远距离拍摄的低分辨率图像中检测脸部关键点, 一直以来都是人脸识别和人机交互领域的难题. 在头部姿势估计的早期研究

中^[22-23], 受限于脸部关键点检测的发展, 基于脸部关键点几何关系的方法一般都假定脸部关键点的位置是已知的, 以便于将精力集中于头部姿势估计模型的研究上. 近年来, 鲁棒脸部关键点检测模型——主动外形模型(Active Appearance Model)的出现为基于脸部关键点几何关系的头部姿势估计方法研究提供新的基础.

脸部关键点几何位置到头部姿势的映射是该类方法的第二个关键步骤. 脸部可用的关键点信息包括眼睛内外角、嘴角、鼻尖、脸部轮廓点等. 不同方法采用的脸部关键点不尽相同, 如文献[22]采用的是标定好的两个外眼角、两个外嘴角及鼻尖, 文献[23]采用的是两个内眼角、两个外眼角及鼻尖, 文献[24]采用的是两个内眼角、两个外眼角及两个嘴角, 文献[28]采用的是两个瞳孔、鼻尖及嘴中线等. 此外, 还有一些更为简单的方法, 仅使用脸部的少量信息, 也取得较好的头部姿势估计效果. 如文献[26]使用脸部的轮廓和肤色信息来估计头部姿势的滚动角分量, 文献[27]通过分析两只眼睛和嘴所组成三角形与等腰三角形的偏离程度来估计头部姿势的滚动角分量和平动角分量.

此类方法中, 较为成功的是文献[29]、[31]的研究成果. 它们利用眼睛、鼻子、嘴巴周围的关键点信息(共 28 个关键点), 设计归一化奇异值分解(Normalized Singular Value Decomposition, NSVD)来估计头部姿势. 其中, 关键点信息是由系统使用检测跟踪的方法自动得到, 实现基于脸部关键点几何关系的自动头部姿势估计系统.

以上方法所采用的脸部关键点是确定的, 如眼角、嘴角、鼻尖等. 文献[30]使用不确定脸部关键点来进行头部姿势估计, 为该类方法提供一种新的研究思路, 但精度有限, 仅能用于粗估计.

总的来说, 基于脸部关键点几何关系的头部姿势估计方法较简单, 时间复杂度较低, 但脸部关键点的检测和定位是其瓶颈所在. 此外, 与基于外观信息的头部姿势估计方法相比, 此类方法对遮挡和关键点的缺失较敏感.

3.4 基于特征回归的方法

头部姿势估计是一个回归问题, 其本质是从图像空间到姿态空间的映射问题. 因此, 一个较自然的思路就是通过机器学习方法, 为头部姿势估计学习训练一个回归函数.

降维是基于特征回归的头部姿势估计方法首先要解决的问题. 数字图像的线性维度一般是较高的, 对于一幅 200×200 的图像, 其维度就达到了 40 000.

数字图像的高维性对于现有回归工具来说是一个极大的挑战. 要应用现有回归工具来实现数字图像到头部姿势的回归, 就需对数字图像进行降维.

降维的思路一般有两种. 一种是使用机器学习或模式识别的基本方法把高维数据直接投影到低维空间. 如文献[33]和[34]使用主成分分析(Principal Component Analysis)来降维, 而文献[35]则使用局部梯度方向直方图(Localized Gradient Orientation Histograms)来降维, 降维之后, 再使用支持向量回归(Support Vector Regression, SVR)来训练回归函数. 实践证明, 上述方法能取得不错的效果. 另一种是提取图像特征的方法. 如文献[36]使用局部梯度方向直方图, 文献[37]、[38]使用 SIFT 特征, 文献[39]使用 LBP 特征, 文献[40]使用 HOG 等特征, 文献[41]使用 Gabor 特征. 另外, 即便是所得特征的维度过高, 也可使用上述第一种方法对其降维.

回归工具的选择是基于特征回归的头部姿势估计方法要解决的第二个关键问题. 尽管可用的回归方法多种多样, 但选择合适的回归工具来进行头部姿势估计还有待进一步研究. 就当前已有头部姿势估计研究成果来看, 神经网络是应用最为广泛的回归工具.

近年来, 随机森林(Random Forest)成为头部姿势研究工作的热点, 各类方法不断涌现. 文献[42]、[43]首次把随机森林引入头部姿势估计. Fanelli 等^[12]使用随机森林作为回归工具, 提出一个基于深度图像的头部姿势估计实时系统, 并在随后的工作中把该方法扩展到 Kinect 的应用上^[13-14]. Tang 等^[15]也使用随机森林, 提出切片模型, 把头部姿势估计的精度推进到一个新台阶. 另外, 其他回归工具, 如偏最小二乘^[44]、谱回归^[45]等在头部姿势估计上也有应用.

总之, 基于特征回归的头部姿势估计是当前主流方法, 具有实时性好、精度高等特点, 但受头部检测定位结果的影响较大.

3.5 基于子空间学习的方法

子空间学习是机器学习和计算机视觉中的一个重要研究课题, 它认为在高维数据中可能存在相关的特征属性和冗余的信息, 可从中学习得到一个低维子空间. 子空间学习的思想为头部姿势估计提供新的研究思路.

在基于子空间学习的头部姿势估计研究中, 流形学习相关理论应用最为广泛. 这是因为, 头部的姿态空间是一个天然的三维空间, 可将其视为嵌入在高维图像空间中的三维姿态流形.

Cheng 等^[46]首次将流形嵌入的思想用来解决头部姿势估计问题, 虽然文献[46]实验所用图像为人工合成, 姿态的变化幅度也较小, 但它揭开流形学习应用于头部姿势估计的新篇章. Raychev 等^[47]使用保距特征映射(ISOMAP)来估计头部姿势, 但实验结果并不理想. 为提高估计精度, 近年来, 文献[48]~[52]基于流形学习的基本思想先后提出新的头部姿势估计方法. 范进富等^[53]综合局部线性嵌入(Locally Linear Embedding, LLE)、局部保持投影(Locality Preserving Projection, LPP)和局部嵌入分析(Locally Linear Embedded, LEA)进行头部姿势估计.

总的来说, 此类方法较新颖, 但目前时间复杂度较高、精度较低, 离实用还有一定距离.

3.6 基于局部约束模型的方法

局部约束模型(Constrained Local Model)是模式识别领域较为重要的目标检测算法, 它把待检测目标划分为若干子目标, 在图像上独立地检测各子目标的位置, 最终通过子目标间位置约束关系确定待检测目标的位置.

人脸是应用局部约束模型思想的天然对象. 人脸是由眼睛、鼻子、嘴、下巴等多个器官组成的, 如图5所示, 各器官的外观不尽相同, 而且它们的相对位置受制于脸部的整体框架, 不随脸部表情的变化而发生改变. 这样, 可为脸部子器官单独训练一个检测器, 然后通过脸部子器官间的拓扑关系来形成脸部模型. 此脸部模型既包含脸部的整体拓扑信息, 又包含各子器官的局部信息, 不受刘海长短、脸型胖瘦等个性化信息的影响, 能较为精确地检测到脸部各部分的位置^[58-59].

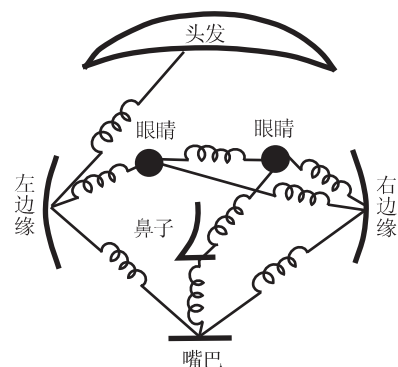


图5 脸部局部约束模型示意图

Fig.5 Graphic description of constrained local model

局部约束模型的思想可用于离散头部姿势估计^[55-56],其基本流程为首先把整个姿态空间划分为若干标准离散姿态,并为每个姿态训练一个脸部模型;新来图像与所有标准姿态的脸部模型进行比较,匹配分数最高脸部模型所对应的姿态即为估计结果。此类方法的流程与基于形状模板的头部姿势估计方法较为类似,不同的地方在于模板模型和匹配方式。基于形状模板的头部姿势估计方法直接使用图像作为模板,新来图像与模板图像在像素层面进行比较;而基于部分模型的头部姿势估计方法则是使用按一定拓扑排列的子器官检测器集作为模板,新来图像与模板之间的匹配是以脸部拓扑为约束,在脸部子器官检测结果候选集上检索最优结果。

4 头部姿态估计数据库及方法精度比较

4.1 性能评价指标

头部姿势估计是一个典型的回归问题,其性能评价指标主要包括平均绝对误差(Mean Absolute Error, MAE)、绝对误差标准差(Standard Deviation of Absolute Error, SDAE)和方向估计精度(Accuracy)。

1) 平均绝对误差。

平均绝对误差是预测精度最为简单的评价指标,主要用来度量预测值跟实际值之间的距离,其计算公式表示如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|,$$

其中, f_i 表示第 i 个样本的预测值, y_i 表示第 i 个样本的真实值。

2) 绝对误差标准差。

平均绝对误差仅能从总体上反映预测值跟实际值之间的距离,不能反映误差的分布情况。绝对误差标准差是各绝对误差偏离平均数距离的平均数,能反映回归误差的离散程度,计算公式如下:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (|f_i - y_i| - MAE)^2}.$$

总的来说,绝对误差标准差是绝对误差平均值分散程度的一种度量。一个较大的标准差,代表大部分绝对误差和其平均值之间差异较大;一个较小的标准差,代表这些绝对误差较接近平均值。

3) 方向估计精度。

如 2.1 节所述,头部姿势估计的一个重要目的是估计脸部朝向。因此,方向估计精度是衡量头部姿

势估计系统的一个重要指标。

要计算头部姿势估计系统的方向估计精度,首先要计算其方向估计误差。方向估计误差为预测所得脸部朝向与实际脸部朝向之间的偏差,如图 6 所示,即为预测所得脸部朝向 $O_p(x_p, y_p, z_p)$ 与实际脸部朝向 $O(x, y, z)$ 之间的夹角 θ :

$$\begin{aligned} \theta &= \arccos(O \cdot O_p) \\ &= \arccos(x \cdot x_p + y \cdot y_p + z \cdot z_p). \end{aligned}$$

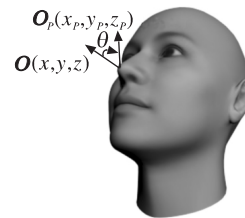


图 6 方向估计误差示意图

Fig. 6 Estimation error of face orientation

对于给定测试集 Ω , 已知方向估计误差 θ , 方向估计精度计算如下:

$$\text{acc}(t) = \frac{N(\Phi)}{N(\Omega)},$$

其中 t 表示误差阈值, $\Phi = \{i \mid \theta < t, i \in \Omega\}$ 表示估计结果的方向误差 θ 在阈值 t 范围内的所有测试样本的集合, $N(\Phi)$ 表示集合 Φ 中的样本个数, $N(\Omega)$ 表示集合 Ω 中的样本个数。

如上所述,头部姿势脸部朝向表达方式是计算方向估计精度的基础。如若系统采用欧拉旋转角作为头部姿势的表达方式,那么就需使用 2.1 节所述方法,将欧拉旋转角转换为对应的脸部朝向,进而计算系统的方向估计精度。

以上 3 种指标中,平均绝对误差和绝对误差标准差是评价回归方法的通用指标,而方向估计精度则是头部姿势估计方法特有的评价指标。一般来说,需依据这 3 个指标对头部姿势估方法进行客观评价。但由于受篇幅限制,一些论文仅从平均绝对误差角度对其方法进行评价。下面以这三项指标为基础,介绍近 5 年头部姿势方法的估计性能。

4.2 主流数据库及方法精度比较

数据库是解决模式识别问题的依托,也是客观评价方法性能的基础。当前的头部姿势估计数据库多种多样,文献[2]详尽描述 2007 年以前创建的头部姿势数据库。近年来,随着深度摄像机技术的发展,头部姿势估计领域出现两个新的深度图像数据库,它们分别是 ETH Face Pose^[9] 和 Biwi Kinect^[40]

数据库. 其中, ETH Face Pose 数据库使用双目深度摄像机搭建, 而 Biwi Kinect 数据库中的图像则是使用 Kinect 深度传感器采集. 表 2 详细介绍这两个新兴数据库及当前常用彩色图像数据库, 并对近 5 年来以这些数据库为实验数据的头部姿势估计方法的精度进行归纳和总结.

表 2 中, 精度栏目中带“*”的为分类结果; 方法类别中的编号对应第 3 节 6 种方法的小节号. 从表 2 中可看出除 BU Dataset 之外, 其他彩色图像数据库中的姿态值都是离散的, 即这些数据库仅包含若

干姿态下的头部图像序列. 以这些数据库为实验数据的头部姿势估计方法大多把头部姿势估计建模成一个分类问题. 然而, 实际上头部姿势空间是一个三维连续空间, 头部姿势估计问题应该是一个回归问题. 因此, 上述彩色头部姿势数据库虽然在一定程度上满足当前研究的需要, 但离现实应用还有一定的距离. 彩色图像数据库 BU Dataset 和深度图像数据库 ETH Face Pose、Biwi Kinect 中的姿态值是连续的, 为研究现实可用的头部姿势估计方法提供数据基础.

表 2 头部姿势估计主流数据库及方法精度比较情况

Table 2 Most widely used head pose databases and performance comparison of corresponding method on these databases

数据库	数据情况	姿态情况	代表方法		误差 /°			精度 /% (<i>thres</i> = 10°)
			方法	类别	Yaw	Pitch	Roll	
Pointing'04	RGB 图像 15 人 186 帧 / 人	$Yaw \in \{-90^\circ, -60^\circ, \dots, 60^\circ, 90^\circ\}$, 9 个离散姿态; $Pitch \in \{-90^\circ, -75^\circ, \dots, 75^\circ, 90^\circ\}$, 13 个离散姿态	文献[37]	3.4	6.05	5.84	—	—
			文献[42]	3.4	—	14.39	—	62.58*
			文献[45]	3.4	13.5	11.3	—	—
CMU-PIE	RGB 图像 68 人 41368 帧图像	$Yaw \in \{-90^\circ, -75^\circ, \dots, 75^\circ, 90^\circ\}$, 13 个离散姿态	文献[51]	3.5	4.33	—	—	—
			文献[44]	3.4	5.31	—	—	79.48*
			文献[56]	3.6	—	—	—	96*
			文献[50]	3.5	14.24	—	—	—
			文献[39]	3.4	—	—	—	98.5*
CAS-PEAL	RGB 图像 1040 人 99594 帧图像	$Yaw \in \{-90^\circ, -67.5^\circ, \dots, 67.5^\circ, 90^\circ\}$, 7 个离散姿态; $Pitch \in \{-45^\circ, 0^\circ, 45^\circ\}$, 3 个离散姿态	文献[39]	3.4	—	—	—	94.5*
			文献[42]	3.4	0.42	—	—	97.23*
FacePix	RGB 图像 30 人 16290 帧图像	$Yaw \in \{-90^\circ, -89^\circ, -88^\circ, \dots, 89^\circ, 90^\circ\}$, 181 个离散姿态	文献[52]	3.5	3.16	—	—	—
			文献[49]	3.5	2.75	—	—	—
			文献[45]	3.4	6.4	—	—	—
			文献[48]	3.5	1.5	—	—	—
			文献[55]	3.5	1.44	—	—	—
BU Dataset	RGB 图像 5 人 45 个图像序列 200 帧 / 序列	连续	文献[32]		3.67	4.97	2.91	—
			文献[1]		5.26	6.1	3.00	—
ETH Face Pose	深度图像 20 人 26 个图像序列 10545 帧图像	连续 $Yaw \in [-90^\circ, 90^\circ]$, $Pitch \in [-45^\circ, 45^\circ]$	文献[9]	3.1	6.1	4.2	—	80.8
			文献[12]	3.4	5.7	5.1	—	90.4
			文献[15]、[18]	3.4	1.5	2.5	—	95.9
			文献[10]	3.1	1.62	2.05	—	90.1
Biwi Kinect	RGB 图像 + 深度图像 20 人 24 个图像序列 约 15000 帧图像	连续 $Yaw \in [-75^\circ, 75^\circ]$, $Pitch \in [-60^\circ, 60^\circ]$, $Roll \in [-50^\circ, 50^\circ]$	文献[13]、[14]	3.4	8.9	8.5	7.9	78.5
			文献[55]	3.6	6.29	5.10	11.29	—
			文献[10]	3.1	1.0	1.1	1.6	91.4

表2中所列头部姿势估计方法为近5年来的新兴头部姿势估计方法. 在估计精度上, 似乎基于彩色图像的头部姿势估计方法的精度高于基于深度图像方法的精度. 如文献[39]为基于彩色图像的头部姿势估计方法, 其精度达到98.5%; 文献[15]、[16]为基于深度图像的头部姿势估计方法, 其 10° 误差以内精度仅为95.9%. 但基于彩色图像头部姿势估计方法的精度为分类精度, 而基于深度图像头部姿势估计方法的估计精度为回归精度. 因此, 两类精度不存在可比性.

本文认为基于深度图像头部姿势估计方法的性能要优于基于深度图像的头部姿势估计方法. 这是因为深度图像受光照变化、表情变化等因素的影响较小, 同时深度图像蕴含的信息具有明显物理意义, 便于研究具备物理意义的表达模型(如文献[15]、[16]). 这也是基于深度图像的头部姿势估计方法成为该领域热点方法的主要原因.

5 结束语

本文结合近年来头部姿势估计问题上的研究工作, 从方法依赖的数据源、人工干预程度和实现原理等三个角度, 综述头部姿势估计问题的研究现状.

从方法的人工干预程度来看, 半自动化的头部姿势估计方法已成为历史, 设计全自动的方法是当前头部姿势估计研究工作的一个基本要求.

从方法的实现原理来看, 基于特征回归的方法具有一定的优越性, 是当前头部姿势估计研究的主流方法. 此类方法采用从图像中提取特征, 训练姿态回归器的方式来实现头部姿势估计. 但当前头部姿势估计领域, 在“什么样的特征能表达头部姿势的本质”、“什么样的头部姿势估计方法最鲁棒”等问题上还未达成共识. 探索头部姿势的本质表达特征仍将是该领域研究的主要课题.

从方法所依赖的数据源来看, 受光照变化、表情变化和遮挡等因素的影响, 基于二维彩色图像的头部姿势估计方法较难取得突破性的发展. 而基于深度图像的方法为头部姿势估计开辟一个新的发展空间. 该类方法不仅能较好克服光照变化的影响, 而且能满足实时性的需要, 成为当前头部姿势估计研究的热点. 但这并不代表着深度信息比彩色信息更加重要, 至少从包含的信息量上来说, 彩色图像要优于深度图像.

融合彩色图像和深度信息(即RGB-D图像)来

解决头部姿势估计问题将成为新的发展趋势. 这是因为使用RGB-D图像来解决视觉问题更加接近人类的视觉原理. 人类视觉系统不仅能感知物理世界的色彩信息, 还能准确感知物理世界中的深度信息, 具备极强的数据感知、内容理解的能力. 这使得人类视觉系统成为计算机视觉无法相比的模仿对象. 但如果计算机视觉系统使用RGB-D图像来作为输入数据, 那么至少使得计算机视觉系统在数据源上与人类视觉系统保持一致.

目前, 头部姿势估计主要应用于两大领域: 生物特征识别领域和人机交互领域. 无论哪种应用, 都需头部姿势估计方法能为其提供实时的头部姿势估计结果. 非实时的头部姿势估计方法常常是无用的. 因此, 实时性是头部姿势估计系统的生命线. 对头部姿势估计系统来说至关重要. 实时的头部姿势估计方法也必将成为头部姿势估计的发展趋势.

另外, 随着头部姿势估计研究的深入发展, 对各种方法进行客观评价也是一项极为重要的工作. 完成这项工作的前提是构建一系列为研究者所公认的测试图像集. 当前头部姿势估计的数据库比较杂乱, 二维彩色图像数据库多达十几种, 较为常用有Pointing'04和CMU PIE, 而深度图像数据库仅有两个. 虽然具有一定规模和涵盖面, 但仍不能满足统一评价各算法的要求. 因此还有待于建立规模更大、覆盖面更广、包含各种类别数据(如彩色、深度图像等)的测试集.

总之, 当前头部姿势估计的方法离高精度、高实时还有一定的距离. 各种头部姿势估计的新方法将不断涌现. 相关研究必将逐步走向成熟. 相关方法必将逐步走向实际应用.

参 考 文 献

- [1] Valenti R, Sebe N, Gevers T. Combining Head Pose and Eye Location Information for Gaze Estimation. *IEEE Trans on Image Processing*, 2012, 21(2): 802-815
- [2] Murphy-Chutorian E, Trivedi M M. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2009, 31(4): 607-626
- [3] Kuipers J B. Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality. Princeton, USA: Princeton University Press, 2002
- [4] Deymer D J. Face Recognition under Varying Pose // *Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 1994: 756-761
- [5] Ng J, Gong S. Composite Support Vector Machines for Detection of Faces across Views and Pose Estimation. *Image and Vision Compu-*

- ting, 2002, 20(5/6): 359–368
- [6] Ng J, Gong S. Multi-View Face Detection and Pose Estimation Using a Composite Support Vector Machine across the View Sphere // Proc of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. Corfu, Greece, 1999: 14–21
- [7] Sherrah J, Gong S, Ong E J. Understanding Pose Discrimination in Similarity Space // Proc of the 10th British Machine Vision Conference. Nottingham, UK, 1999: 523–532
- [8] Sherrah J, Gong S, Ong E J. Face Distributions in Similarity Space under Varying Head Pose. Image and Vision Computing, 2001, 19(12): 807–819
- [9] Breitenstein M D, Kuettel D, Weise T, *et al.* Real-Time Face Pose Estimation from Single Range Images. // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008. DOI: 10.1109/CVPR. 2008.4587807
- [10] Padelaris P, Zabulis X, Argyros A A. Head Pose Estimation on Depth Data Based on Particle Swarm Optimization // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Providence, USA, 2012: 42–49
- [11] Malassiotis S, Srinivas M G. Robust Real-Time 3D Head Pose Estimation from Range Data. Pattern Recognition, 2005, 38(8): 1153–1165
- [12] Fanelli G, Gall J, Van Gool L. Real Time Head Pose Estimation with Random Regression Forests // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 617–624
- [13] Fanelli G, Weise T, Gall J, *et al.* Real Time Head Pose Estimation from Consumer Depth Cameras // Proc of the 33rd DAGM Symposium. Frankfurt, Germany, 2011: 101–110
- [14] Fanelli G, Dantone M, Gall J, *et al.* Random Forests for Real Time 3D Face Analysis. International Journal of Computer Vision, 2013, 101(3): 437–458
- [15] Tang Y Q, Sun Z N, Tan T N. Face Pose Estimation Based on Integral Slice Features of Single Depth Images // Proc of the 1st Asian Conference on Pattern Recognition. Beijing, China, 2011: 530–534
- [16] Tang Y Q, Sun Z N, Tan T N. Real-Time Head Pose Estimation Using Random Regression Forests // Proc of the 6th Chinese Conference on Biometric Recognition. Beijing, China, 2011: 66–73
- [17] Seemann E, Nickel K, Stiefelhofen R. Head Pose Estimation Using Stereo Vision for Human-Robot Interaction // Proc of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Seoul, Republic of Korea, 2004: 626–631
- [18] Wu J W, Trivedi M M. A Two-Stage Head Pose Estimation Framework and Evaluation. Pattern Recognition, 2008, 41(3): 1138–1158
- [19] Gurbuz S, Oztup E, Inoue N. Model Free Head Pose Estimation Using Stereovision. Pattern Recognition, 2012, 45(1): 33–42
- [20] Rowley H A, Baluja S, Kanade T. Rotation Invariant Neural Network-Based Face Detection // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Santa Barbara, USA, 1998: 38–44
- [21] Wilson H R, Wilkinson F, Lin L M, *et al.* Perception of Head Orientation. Vision Research, 2000, 40(5): 459–472
- [22] Gee A, Cipolla R. Determining the Gaze of Faces in Images. Image and Vision Computing, 1994, 12(10): 639–647
- [23] Horprasert T, Yacoob Y, Davis L S. Computing 3D Head Orientation from a Monocular Image Sequence // Proc of the 2nd International Conference on Automatic Face and Gesture Recognition. Killington, USA, 1996: 242–247
- [24] Wang J G, Sung E. EM Enhancement of 3D Head Pose Estimated by Point at Infinity. Image and Vision Computing, 2007, 25(12): 1864–1874
- [25] Huang J, Shao X H, Wechsler H. Face Pose Discrimination Using Support Vector Machines // Proc of the 14th International Conference on Pattern Recognition. Brisbane, Australia, 1998: 154–156
- [26] Cordea M D, Petriu E M, Georganos N D, *et al.* Real-Time 2(1/2)-D Head Pose Recovery for Model-Based Video-Coding. IEEE Trans on Instrumentation and Measurement, 2001, 50(4): 1007–1013
- [27] Nikolaidis A, Pitas I. Facial Feature Extraction and Pose Determination. Pattern Recognition, 2000, 33(11): 1783–1791
- [28] Liang X N, Tong W Q. Face Pose Estimation Using Near-Infrared Images // Proc of the International Conference on Communication Systems and Network Technologies. Rajkot, India, 2012: 216–220
- [29] Zhu Z W, Ji Q. Robust Real-Time Face Pose and Facial Expression Recovery // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 681–688
- [30] Xu T, Wang C, Wang Y H, *et al.* Saliency Model Based Head Pose Estimation by Sparse Optical Flow // Proc of the 1st Asian Conference on Pattern Recognition. Beijing, China, 2011: 575–579
- [31] Tong Y, Wang Y, Zhu Z W, *et al.* Robust Facial Feature Tracking under Varying Face Pose and Facial Expression. Pattern Recognition, 2007, 40(11): 3195–3208
- [32] Morency L P, Whitehill J, Movellan J. Monocular Head Pose Estimation Using Generalized Adaptive View-Based Appearance Model. Image and Vision Computing, 2010, 28(5): 754–761
- [33] Li Y G, Gong S, Liddell H. Support Vector Regression and Classification Based Multi-View Face Detection and Recognition // Proc of the 4th IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble, France, 2000: 300–305
- [34] Li Y M, Gong S, Sherrah J, *et al.* Support Vector Machine Based Multiview Face Detection and Recognition. Image and Vision Computing, 2004, 22(5): 413–427
- [35] Murphy-Chutorian E, Doshi A, Trivedi M M. Head Pose Estimation for Driver Assistance Systems: A Robust Algorithm and Experimental Evaluation // Proc of the IEEE Conference on Intelligent Transportation Systems. Seattle, USA, 2007: 709–714
- [36] Murphy-Chutorian E, Trivedi M M. Head Pose Estimation and Augmented Reality Tracking: An Integrated System and Evaluation for Monitoring Driver Awareness. IEEE Trans on Intelligent Transportation Systems, 2010, 11(2): 300–311

- [37] Ho H T, Chellappa R. Automatic Head Pose Estimation Using Randomly Projected Dense SIFT Descriptors // Proc of the IEEE International Conference on Image Processing. Orlando, USA, 2012: 153–156
- [38] Chen C, Schonfeld D. A Particle Filtering Framework for Joint Video Tracking and Pose Estimation. IEEE Trans on Image Processing, 2010, 19(6): 1625–1634
- [39] Ma B P, Chai X J, Wang T J. A Novel Feature Descriptor Based on Biologically Inspired Feature for Head Pose Estimation. Neurocomputing, 2013, 115(4): 1–10
- [40] Dong L G, Tao L M, Xu G Y. Head Pose Estimation Based on a Second Order Histogram of the Orientation Gradient. Journal of Tsinghua University: Science & Technology, 2011, 51(1): 73–79 (in Chinese)
(董力庚, 陶霖密, 徐光祐. 基于二阶梯度朝向直方图特征的头部姿态估计. 清华大学学报: 自然科学版, 2011, 51(1): 73–79)
- [41] Ma B P, Shan S G, Chen X L, et al. Robust Appearance-Based Method for Head Pose Estimation. Journal of Software, 2009, 20(6): 1651–1663 (in Chinese)
(马丙鹏, 山世光, 陈熙霖, 等. 一种基于图像表观的鲁棒姿态估计方法. 软件学报, 2009, 20(6): 1651–1663)
- [42] Huang C, Ding X Q, Fang C. Head Pose Estimation Based on Random Forests for Multiclass Classification // Proc of the 20th International Conference on Pattern Recognition. Istanbul, Turkey, 2010: 934–937
- [43] Li Y L, Wang S G, Ding X Q. Person-Independent Head Pose Estimation Based on Random Forest Regression // Proc of the 17th IEEE International Conference on Image Processing. Hong Kong, China, 2010: 1521–1524
- [44] Haj M A, Gonzalez J, Davis L S. On Partial Least Squares in Head Pose Estimation: How to Simultaneously Deal with Misalignment // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2602–2609
- [45] Shan C F, Chen W. Head Pose Estimation Using Spectral Regression Discriminant Analysis // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 116–123
- [46] Chen L, Zhang L, Hu Y X, et al. Head Pose Estimation Using Fisher Manifold Learning // Proc of the International Workshop on Analysis and Modeling of Faces and Gesture. Nice, France, 2003: 203–207
- [47] Raytchev B, Yoda I, Sakaue K. Head Pose Estimation by Nonlinear Manifold Learning // Proc of the 17th International Conference on Pattern Recognition. Tsukuba, Japan, 2004, IV: 462–466
- [48] Ben Abdelkader C. Robust Head Pose Estimation Using Supervised Manifold Learning // Proc of the 11th European Conference on Computer Vision. Heraklion, Greece, 2010: 518–531
- [49] Foytik J, Asari V K. A Two-Layer Framework for Piecewise Linear Manifold-Based Head Pose Estimation. International Journal of Computer Vision, 2013, 101(2): 270–287
- [50] Lu J W, Tan Y P. Ordinary Preserving Manifold Analysis for Human Age and Head Pose Estimation. IEEE Trans on Human-Machine Systems, 2013, 43(2): 249–258
- [51] Huang D, Storer M, de la Torre F, et al. Supervised Local Subspace Learning for Continuous Head Pose Estimation // Proc of the IEEE Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 2921–2928
- [52] Liu X Y, Lu H T, Zhang D Q. Head Pose Estimation Based on Manifold Embedding and Distance Metric Learning // Proc of the 9th Asian Conference on Computer Vision. Xi'an, China, 2009: 61–70
- [53] Fan J F, Chen D S. Combining Manifold Learning and Nonlinear Regression for Head Pose Estimation. Journal of Image and Graphics, 2012, 17(8): 1002–1010 (in Chinese)
(范进富, 陈锻生. 流形学习与非线性回归结合的头部姿态估计. 中国图象图形学报, 2012, 17(8): 1002–1010)
- [54] Wang X W, Huang X Y, Gao J Z, et al. Illumination and Person-Insensitive Head Pose Estimation Using Distance Metric Learning // Proc of the 10th European Conference on Computer Vision. Marseille, France, 2008: 624–637
- [55] Baltrusaitis T, Robinson P, Morency L. 3D Constrained Local Model for Rigid and Non-Rigid Facial Tracking // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2610–2617
- [56] Zhu X X, Ramanan D. Face Detection, Pose Estimation, and Landmark Localization in the Wild // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 2879–2886
- [57] Viola P, Jones M. Rapid Object Detection Using a Boosted Cascade of Simple Features // Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, USA, 2001: 511–518
- [58] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing Parts of Faces Using A Consensus of Exemplars // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA, 2011: 545–552
- [59] Belhumeur P N, Jacobs D W, Kriegman D J, et al. Localizing Parts of Faces Using a Consensus of Exemplars. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2930–2940