

基于树结构分层随机森林在非约束环境下的头部姿态估计

刘袁缘 陈靓影^{*} 俞侃 覃杰 陈超原

(文华学院 武汉 430074)

(华中师范大学国家数字化学习工程研究中心 武汉 430079)

(武汉华中数控股份有限公司红外事业部 430074)

摘要: 头部姿态估计是人类行为和注意力的关键,受到光照、噪声、身份、遮挡等许多因素的影响。为了提高非约束环境下的估计准确率和鲁棒性,该论文提出了树结构分层随机森林在非约束环境下的多类头部姿态估计。首先,为了消除不同环境的噪声影响,提取人脸区域的组合纹理特征,对人脸区域进行积极人脸子区域的分类,分类结果作为树结构分层随机森林的先验知识输入;其次,提出了一种树结构分层随机森林算法,分层估计多自由度下的头部姿态;再次,为了增强算法的分类能力,使用自适应高斯混合模型作为多层次子森林叶子节点的投票模型。在多个公共数据集上的多种非约束实验环境下进行头部姿态估计,最终实验结果表明所提算法在不同质量的图像上都有很好的估计准确率和鲁棒性。

关键词: 头部姿态估计;非约束环境;树结构分层随机森林;人脸积极子区域先验分类;自适应高斯混合模型

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2015)03-0543-09

DOI: 10.11999/JEIT140433

Head Pose Estimation Based on Tree-structure Cascaded Random Forests in Unconstrained Environment

Liu Yuan-yuan Chen Jing-ying Yu Kan Qin Jie Chen Chao-yuan

(Wenhua College, Wuhan 430074, China)

(National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China)

(Wuhan Huazhong Numerical Control Co., Ltd, Wuhan 430074, China)

Abstract: Head pose estimation is an important evaluating indicator of human attention, which depends on many factors, such as illumination, noise, identification, occlusion and so on. In order to enhance estimation efficiency and accuracy, this paper presents tree-structure cascaded random forests to estimate head pose in different quality images. First, in order to eliminate the influence of different environment noise, combined texture features in random forests for positive facial patch classification are extracted, which will be the privileged inputs to estimate head pose. Second, a coarse-to-fine approach is proposed to estimate head pose both in the yaw and pitch, which is called tree-structure cascaded random forests. Third, an adaptive Gaussian mixture model is used to enhance discriminate vote energy in the tree distribution. This framework is evaluated in unconstrained environmental datasets. The experiments show that the proposed approach has a remarkable and robust performance in different quality images.

Key words: Head pose estimation; Unconstrained environment; Tree-structure cascaded random forests; Positive facial patch privileged classification; Adaptive Gaussian mixture model

1 引言

头部姿态是研究人类行为和注意力的关键^[1]。因此,在许多智能系统中非约束环境下的头部姿态估计是检测人类身份和行为的重要环节。但是非约束环境中的投影几何形变、背景光照变化、前景遮挡

问题和低分辨率等因素的影响,使得头部姿态的多自由度估计一直是一个富有挑战性的领域^[2]。已有的一些头部姿态估计方法根据特征选择的不同,大致可以分为基于局部特征的方法和基于全局处理的方法。前者依赖于局部点特征的提取,后者则是对整个人脸区域进行处理。基于局部特征的方法通常先提取人脸的特征点,如眼睛点、眉毛点以及嘴角等,它主要适用于高精度的系统和人脸的高分辨率图像^[3]。对于非约束的环境下,有很强的局限性。而基

2014-04-13 收到, 2014-07-25 改回

国家自然科学基金(61205062),湖北省自然科学基金(2012FFB02701)和华中科技大学文华学院青年基金(J0200540102)资助课题

*通信作者: 陈靓影 chenjy@mail.ccnu.edu.cn

于全局的方法的优势是只需要处理整个人脸的图像区域而不需要提取局部的点特征,适用于低质量低分辨率的图像估计。基于全局的方法主要是用机器学习和模式识别的算法,如模板匹配算法、多分类器阵列算法^[4]、SVM^[5]、随机森林^[6]、Adaboost 等。

随机森林^[7]因为具有快速处理大数据的训练能力和高效的在线计算能力,成为近年来机器视觉中处理大数据量的热门方法之一。最近,随机森林已经应用于实时的 2D 头部姿态的估计和分类^[8]以及 3D 图像的头部姿态识别^[9-11]。文献[6]提出了条件随机森林来检测 2D 头部姿态估计和人脸特征点。其在水平自由度的 5 个旋转角度可达到识别率 72.3%。文献[4]使用组合回归和分类随机森林的算法,获得了较好的估计效果。但是对环境的要求都有一定的限制性。为了在非约束环境下进行多类头部姿态的准确估计,本文提出了树结构分层随机森林算法,用于非约束环境下头部姿态的多自由度的鲁棒估计。

树结构的分层概率模型是文献[12]提出的,已经证明了它的高效率和高准确率,许多文献也已经用它进行多目标跟踪和情感计算。树结构的分层概率模型每一层节点概率是其上一层节点概率和分支选择概率的结果,即每一个子层都收到其父层的影响。每次计算树的概率时,只需要用相关分支上的节点概率,而不需要计算整棵树所有节点的概率。因此引入树结构分层随机森林具有更高的效率和准确率。

本文的创新点如下:(1)积极人脸子区域的先验分类可以消除不同环境下的噪声影响;(2)一种树结构的多层随机森林算法的提出,提高了非约束环境下多类头部姿态估计的准确率和效率;(3)自适应高斯混合模型作为多层次子森林叶子节点的投票模型使得分类结果具有更强的鲁棒性。

2 基于树结构分层随机森林对头部姿态的估计

为了更好地在非约束环境下对多自由度的头部姿态进行估计,本文提出了基于树结构分层随机森

林的多层估计算法。算法如图 1 所示,分为 3 个阶段。第 1 阶段,为了消除非约束复杂背景的干扰,我们提取了人脸积极的子区域块作为分层随机森林的先验输入;第 2 个阶段本文提出树结构分层随机森林来估计水平头部姿态,级联分布的两个子层次为 S-1 和 S-2;第 3 个阶段在水平头部姿态估计结果的条件下进行竖直方向的头部姿态估计,其中级联分布的两个子层为 S-3 和 S-4。最终在 S-4 子层得到 25 对头部姿态的估计结果。

2.1 人脸子区域的先验分类

对于头部姿态估计有两个难点,一个是正确提取人脸区域,一个是非约束背景的影响。在非约束背景中提取人脸的前景目标区域,首先要去除背景信息的干扰,包括头发、遮挡、背景、光照等。因此,我们将人脸子区域分为两个子区域:人脸积极子区域和干扰子区域。人脸积极子区域是去除噪声的区域,对头部姿态的估计有积极的影响,反之则为干扰区域,如图 2 所示,并将分类结果作为头部姿态估计的先验知识。

为了去除干扰子区域块对头部姿态的影响,我们用 Haar 特征检测到的人脸区域进行人脸积极子区域块的先验分类。如图 3 所示。首先,对人脸区域随机提取 200 个子区域块,并提取每个子区域块的 Gabor 特征符。然后,我们用分类随机森林^[4]离线训练人脸积极子区域和干扰子区域类,并用 1 和 0 分别进行标注。分类随机森林的计算过程是每颗决策树迭代的过程,森林中的每一棵树 T 都是由标注好的随机提取的数据集训练生成。当测试数据 P 通过树的根节点到达其叶子节点时,存储在叶子结点的概率密度 $p(c = k | l_i(P))$ 将判断测试数据 P 的类别信息。对于每个叶子结点的子区域块的概率直接作为头部姿态估计的先验输入,仅仅对分类结果为 1 的人脸积极子区域块进行头部姿态的估计。

2.2 树结构分层随机森林的训练

树结构是一种级联式的多层概率分布模型,它的当前叶子节点概率 $[p_1, \dots, p_i]$ 是它上一层节点概率的 $[a_1, a_2, \dots, a_k]$ 在其相关分支 b_{ji} 上的结果^[9]。图 4 中,

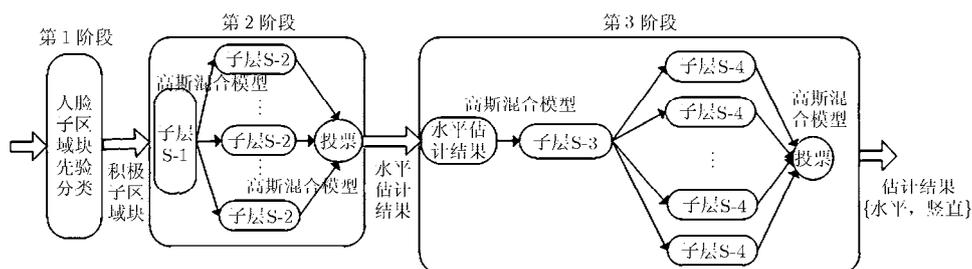


图 1 基于树结构分层随机森林的头部姿态估计算法流程图

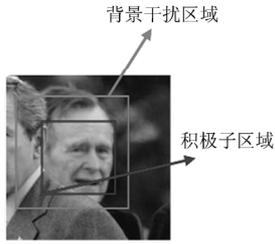


图 2 人脸积极子区域和背景干扰区域

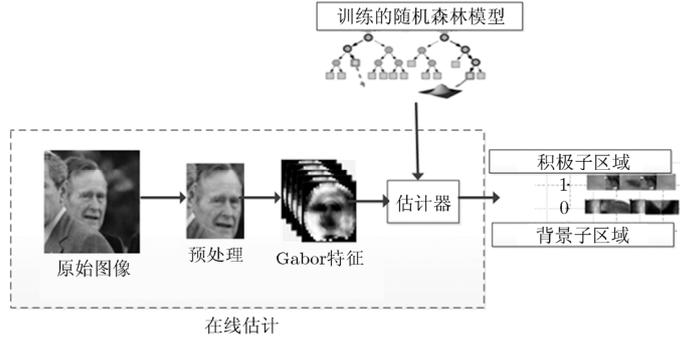


图 3 人脸积极子区域的先验分类

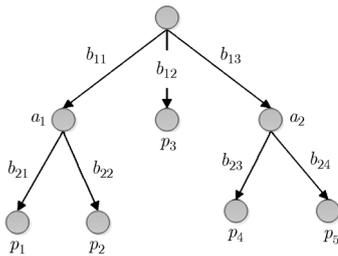


图 4 树结构分布的多层概率模型

i 表示叶子节点的序号, k 是它上一层节点的序号, j 表示该分支号。可见在树结构分布的概率模型中, 每一个子层只与它的父层有关。因此, 树结构分布的多层概率模型中, 只需要计算子层中当前子树的概率模型和它父层的先验概率模型, 而不需要计算随机森林中所有树的概率模型。所以, 树结构的分层概率模型可以提供更好的准确率和效率。

为了构建树结构分层随机森林来估计多自由度的头部姿态, 我们需要:

- (1) 标注满足树结构级联分层的训练数据集 $P_i | a_j$;
- (2) 定义二进制测试 ϕ ;
- (3) 定义级联式树结构随机森林的测度 $H(P | a_j)$;
- (4) 定义存储在叶子节点的自适应投票模型。

级连标注 训练的过程是监督的, 森林中每一棵树 T 的建立都是在不同的数据集中随机训练而成 $T = \{T_i\}$ 。对每一张人脸图像, 我们随机提取人脸子区域块集的组合特征: $P_i = \{I_i^1, I_i^2, I_i^n\}$ 。其中, I_i^1 表示 Gabor 特征, 它的空间维度是 $35 \times 31 \times 31$ 。 I_i^2 为原始的灰度值, 它的空间维度是 31×31 。 c_i 表示头部姿态类的标注:

$$C_i^n = (c_i^1, (c_i^2 | c_i^1), (c_i^3 | c_i^2, c_i^1), (c_i^4 | c_i^3, c_i^2, c_i^1)) \quad (1)$$

c_i^1 是水平自由度旋转的3类头部姿态, $c_i^2 | c_i^1$ 是级联第2层在 c_i^1 条件下的5个水平头部姿态类, $c_i^3 | c_i^2, c_i^1$ 是级联第3层在水平自由度旋转条件下的15对头部

姿态类, $c_i^4 | c_i^3, c_i^2, c_i^1$ 是储存在级联条件随机森林第4层中的水平、竖直双自由度下的25类头部姿态类。

二进制测试 二进制测试是在子集中不断地逼近类标注不确定性纯度的过程, 最终将训练集分裂成2个子集。随机树的生长就是一个由二进制测试创建子节点的迭代过程。我们定义二进制测试 ϕ 为

$$|R_1|^{-1} \sum_{j \in R_1} I^f(j) - |R_2|^{-1} \sum_{j \in R_2} I^f(j) > \tau \quad (2)$$

其中, R_1, R_2 是人脸子区域中的两个随机选取的矩形子块, $I^f(j)$ 是上一部分定义好的特征通道, τ 是阈值。开始测试, 当测试结果大于 τ 时, 生成右子节点, 反之生成左子节点。

树状条件测度 $H(P | a_j)$ 在这部分, 测度 $H(P | a_j)$ 定义为连续子区域的熵。

$$H(P | a_j) = - \sum_{i=1}^N \frac{\sum_i p(c_i | a_j, P_n)}{|P|} \lg \left(\frac{\sum_i p(c_i | a_j, P_n)}{|P|} \right) \quad (3)$$

其中 $p(c_i | a_j, P_n)$ 表示人脸子区域块 P_n 在分层随机森林的第 j 层第 a_j 子森林中属于头部姿态类 c_i 的概率, $|P|$ 是头部姿态类为 c_i 的人脸子区域块的数量。选择最佳分裂申请, 它可以使得信息增益(IG)估计函数最大

$$IG = \arg \max_{\phi} (H(P | a_j) - (\omega_L H(P_L | a_j) + \omega_R H(P_R | a_j))) \quad (4)$$

w_L, w_R 是数据集 P_L (通过上述二进制测试到达左子集的数量) P_R (通过上述二进制测试到达右子集的数量) 的样本数量和总数据集 P 的比率。

叶子 如果信息增益(IG)低于预先设定的阈值或者树的最大深度达到时, 生产一个叶子节点。在每一个叶子结点中, 包括了头部姿态的分类概率和连续头部姿态分布参数, 其满足一个高斯概率分布模型。

$$p(c_i^m | l_{a_j}) = N \left(c_i^m | l_{a_j}; \overline{c_i^m | a_j}, \sum_{a_j} \right) \quad (5)$$

其中, $c_i^m|a_j$ 和 \sum_{a_j} 是树结构分层随机森林的第 j 层中子森林 a_j 的头部姿态概率均值和协方差。

当一个子区域块到达子森林的叶子节点时, 我们用类决策模型 $C(P)$ 加载下一个子森林树。

$$C(P) = \arg \max_{a_j \in C_i^n} p(c_i|a_j, P) \quad (6)$$

其中, $p(c_i|a_j, P)$ 是森林中的第 j 层中子森林 a_j 的条件下的估计概率, 它由下一节所述的自适应高斯混合模型计算得到。最终的头部姿态由自适应高斯混合模型进行投票分类。

自适应高斯混合模型 存储在叶子的概率 $p(c = k | P)$ 具有判断测试子区域块属于头部姿态类 k 的信息。随机森林的叶子结点 l 存储符合多项式高斯分布。

$$p(c_i|l) = N\left[c_i; \bar{c}_i, \sum_l\right], c_i \in \{-90^\circ, -45^\circ, 0^\circ, 45^\circ, 90^\circ\} \quad (7)$$

其中, \bar{c}_i 和 \sum_l 是第 i 类头部姿态的均值和方差。

树结构分层随机森林是在每一个分支层上有选择性的级联子森林。当到达下一分支层时, 它需要做出判断加载哪一颗子森林。因为在不同水平角度下竖直角度的估计投票仍满足高斯模型分布, 所以我们提出一个自适应高斯混合模型来投票最终的头部姿态。改进公式 $C(P)$ 得到:

$$p(c_i | a_j, l_{ji}) = N\left[c_i | a_j, \bar{c}_i | a_j, \sum_{l_{ji}}\right], \quad c_i | a_j = \{\delta_{ji}(k) c_i^j(k)\}, k = 1, 2, 3, 4 \quad (8)$$

i 是树结构分层随机森林的子分支, j 是分支 i 的子节点, k 是子区域块达到的叶子结点中存储的标注姿态。

2.3 基于树结构分层随机森林的水平头部姿态估计

由于水平自由度旋转包含更多的头部姿态信息, 如眼睛、鼻尖和嘴角的信息, 因此我们将其作为树结构分层随机森林估计头部姿态的第1层和第2层, 如图5所示。如2.2节训练树结构分层随机森林的子森林。首先, 量化训练数据在水平子层S-1和S-2的相关头部姿态子集“左”, “正面”, “右”和“正左”, “左中”, “正面”, “右中”, “正右”, 并分别用标注“-1, 0, 1”和“-2, -1, 0, 1, 2”代替真实的头部姿态旋转角度 $-90^\circ \sim 90^\circ$ 。然后将估计结果(水平旋转角度) a 作为竖直估计的父层概率模型 $p(c_i | a)$ 。

2.4 基于树结构分层随机森林的竖直头部姿态估计

树结构分层随机森林级联第3层S-3和第4层S-4, 在水平自由度的估计条件下进行竖直自由度的头部姿态估计。由于缺少更多的人脸信息, 竖直自由度估计是头部姿态估计领域的一个难点。文献[8]

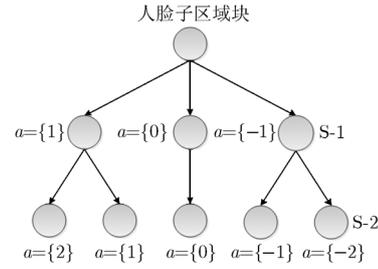


图5 水平方向的分层头部姿态估计

用随机森林同时训练和测试水平和竖直自由度的头部姿态。本文则提出了一个树结构分层随机森林方法来分层估计水平和竖直多自由度的头部姿态。我们将水平自由度的估计结果作为竖直自由度的估计条件输入, 然后对每一个分支树进行3类竖直头部姿态估计, 最后对左子类 and 右子类再次进行细化估计角度, 最终估计出5类竖直头部姿态。级联细化算法的流程结构如图6所示。其中 a 为水平方向的估计结果, 级联分支估计为竖直估计中3类角度的粗糙估计结果, 最终细化估计为竖直估计中的最终细化估计结果。由于2.3节中水平估计的结果为5类水平旋转角度, 因此图6由5棵相同的独立子森林构成。在这个阶段, 我们最终可以检测25个离散的头部运动角度, 检测结果表示为 $\{90^\circ, 90^\circ\}, \{90^\circ, 45^\circ\}, \dots, \{0^\circ, 0^\circ\}, \dots, \{-45^\circ, -90^\circ\}, \{-90^\circ, -90^\circ\}$ 。

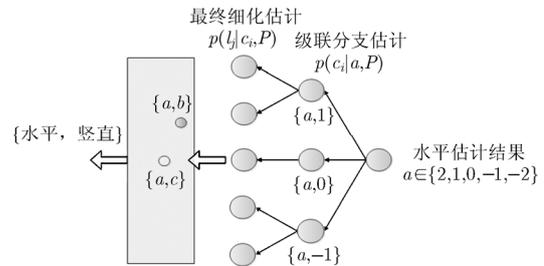


图6 竖直方向的分层头部姿态估计

3 树结构分层随机森林的多层概率模型

随机森林的目标是通过叶子结点构建人脸子区域块 P 的类概率估计 $p(c_i | P)$ [6], 树结构分层随机森林模型的目标则是构建树结构分层的先验条件概率估计 $p(c_i | \alpha, P)$, 本文改进随机森林的概率模型为

$$p(c_i | \alpha, P) = \int \frac{p(c_i, \alpha | P) \cdot p(c_i | P)}{\sum_{i=1}^n p(c_i, \alpha | P) p(c_i | P)} d\alpha \quad (9)$$

式中, α 是上一层估计的概率结果。

为了学习 $p(c_i, \alpha | P)$, 训练集 α 被分裂为不相交的离散子集 a_j 。因此, 式(6)可改写为

$$\begin{aligned}
 & p(c_i | \alpha, P) \\
 &= \sum_j p(c_i, a_j | P) \int \frac{p(c_i | P)}{\sum_{i=1}^n p(c_i, \alpha | P) p(c_i | P)} d\alpha \\
 &= \sum_j \sum_i p(c_i | a_j, p) p(a_j | P) \\
 &\quad \cdot \int \frac{p(c_i | P)}{\sum_{i=1}^n p(c_i, \alpha | P) p(c_i | P)} d\alpha \\
 &= \sum_j \sum_i p(c_i | a_j, p) p(a_j | P) \int \frac{1}{p(\alpha | P)} d\alpha \quad (10)
 \end{aligned}$$

先验概率 $p(c_i | a_j, p)$ 可以在每一个训练子集 a_j 中用改层的子随机森林 $T(a_j)$ 学习得到。同样地，概率 $p(\alpha | P)$ 可以在所有的训练集 a 上用随机森林学习得到。最终，得到不同分支上的多层概率模型为

$$p(c_i | \alpha, P) = \frac{1}{T_t} \sum_j \sum_{t=1}^{k_j} p(c_i | l_{t,a_j}(P)) \quad (11)$$

其中 l_{t,a_j} 表示子区域块 P 在树 $T_t \in T(a_j)$ 中到达的叶子结点。离散值 k_j 由 $\sum_j k_j = T_t$ 和式(12)计算得到。

$$\sum_j k_j \approx T_t \int_{\alpha \in a_j} P(\alpha | P) d\alpha \quad (12)$$

4 实验

为了测试在非约束环境下的估计结果，我们在 Pointing'04 头部姿态数据库^[13]，LFW 数据库^[14]以及

实验室实时采集的数据集上测试本文的方法。Pointing'04 头部姿态数据库是包括 15 个人的两种不同表情的头部姿态数据库，共有 2790 张图片。LFW 数据库包含 5749 个不同个体的人脸图片，这些图片都是自然状态下收集的，包括不同的姿态、光照、分辨率、质量、表情、性别、种族等。我们实验室的实时数据集收集了 20 个不同人的不同姿态、表情、遮挡的图片，其中包括 10 个男性，10 个女性，每个人包括 25 个头部姿态，总共 500 张图片。本文数据集的标注方法参考 LFW 数据库^[14]的标注方法。实验过程中，数据集被分为训练数据集和测试数据集。训练集采用 Pointing'04 数据库中的 2100 张图片，LFW 中 12000 张图片以及实验室数据库的 300 张图片。测试集包括 Pointing'04 数据库中剩余的 690 张图片，LFW 数据库的 1500 张图片和实验室采集的实时数据集的 200 张图片。所有图片都是在包括遮挡、低分辨率、性别、背景干扰、种族等非约束的环境中采集的，估计结果如图 7 所示。第 1 行为加入噪声遮挡的 Pointing'04 头部姿态数据库的估计结果，第 2 行为自然环境下实验室数据集的估计结果，第 3 行为 LFW 数据库的估计结果。估计结果实时地显示在图片中，实验结果表明本文方法对于非约束环境下具有更好的鲁棒性。

4.1 训练

为了训练，需要对随机森林的一些预设参数进行分析。图 8 描述了树的数量、最大深度和分裂次



图 7 本文方法在非约束环境下的估计结果

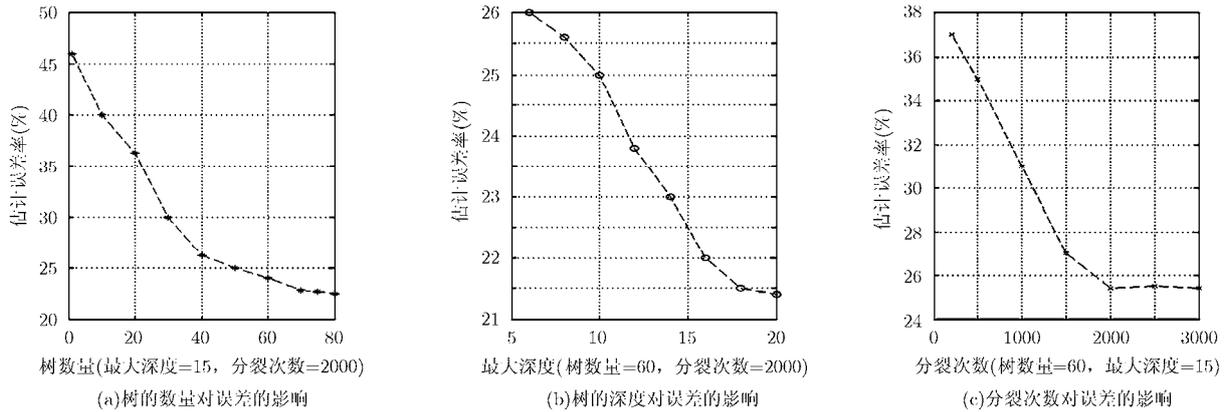


图8 树的数量、最大深度、分裂次数对估计结果的误差分析

数对估计结果的误差分析曲线。估计误差随着树数量的增加，树的最大深度增加以及分裂次数的增加而减少，当参数值增加到一定程度，对误差估计率的影响很小。因此，本文选择树的最大深度为 15，每一个节点的随机分裂次数是 2000，分裂阈值为 25，人脸大小归一化为 125×125 ，人脸子区域块大小为 30×30 。训练时，首先以分层结构的方式从所有的数据集中选择子集，每 186 张图片组成一个子集训练成一棵树。树结构分层森林包括 4 层，第 1 层水平自由度有 15 颗随机树，第 2 层水平自由度有 10 颗随机树，第 3 层垂直自由度有 15 颗随机树，最后一层垂直自由度有 25 颗随机树。且每一层的训练都是上一层条件分支的结果。

4.2 测试

测试数据集包括 Pointing'04 头部姿态数据库中用于训练的 840 张图片，1500 张 LFW 数据库图片，以及 200 实验室数据集。测试参数选择包括随机森林的参数(与训练保持一致)、子森林的树分支数，自适应高斯混合模型参数。由于图像 Gabor 特征包含了方向信息，对于光照和旋转具有很好的鲁棒性。我们从测试图片中稠密提取 200 个人脸积极子区域，并提取子区域的 Gabor 特征进行测试。

由于树结构分层随机森林实际上是将随机森林

以树状结构条件模型进行重新分布，因此它具有更高的搜索决策效率和准确率。为了比较本文方法与随机森林的估计能力，图9描述了随机森林进行头部姿态估计的所有的估计概率投票分布，图10~图12描述了本文方法对头部姿态的估计概率分布。如图所示，随机森林的概率分布在不同的姿态类上有高度的重合，而本文方法在最终的概率分布上基本没有重合。可见，本文方法提高了时间和空间资源的利用率，同时消除了水平自由度旋转和垂直自由度旋转的干扰，具有更强的区分力和分类能力。

4.3 估计准确率比较和分析

为了更好地比较树结构分层随机森林和随机森林算法，两个算法的测试过程选择相同的测试图片，相同的特征参数。实验结果如表1所示，其中列D描述树结构分层随机森林算法的估计准确率，列R描述随机算法的估计准确率。每个方格均为离散的不重复的 $45^\circ \times 45^\circ$ 区域。树结构分层随机森林算法的平均准确率可以达到 71.83%，而随机森林只能达到 62.23%。

4.4 树结构分层随机森林的级联层数分析

图13显示估计准确率与级联层数的关系，S-0表示为1层树分布的25类头部姿态估计准确率为 62.23%，S-2为2层树分布的估计准确率上升为

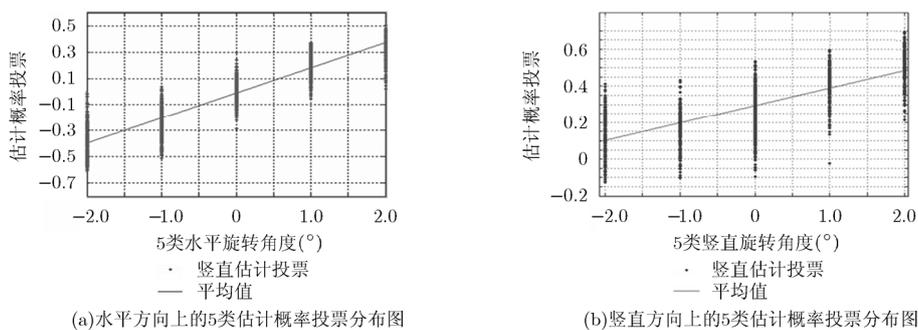


图 9 随机森林不同姿态类上的估计概率投票分布

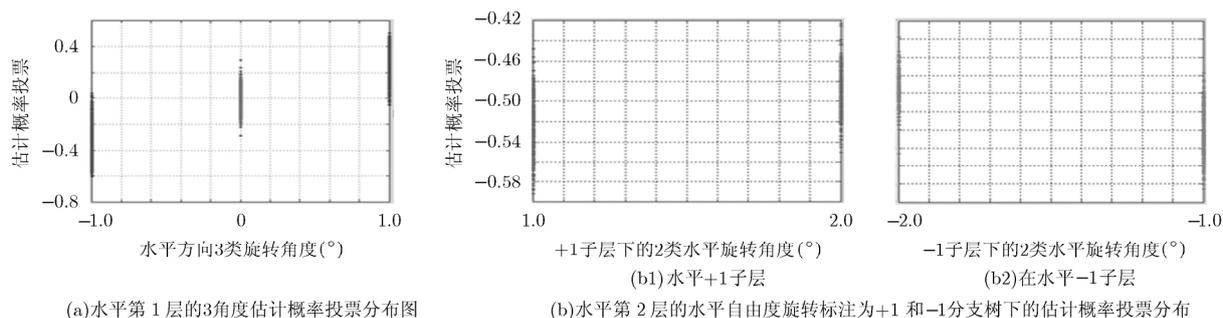


图 10 树结构分层随机森林在水平方向上的估计概率投票分布

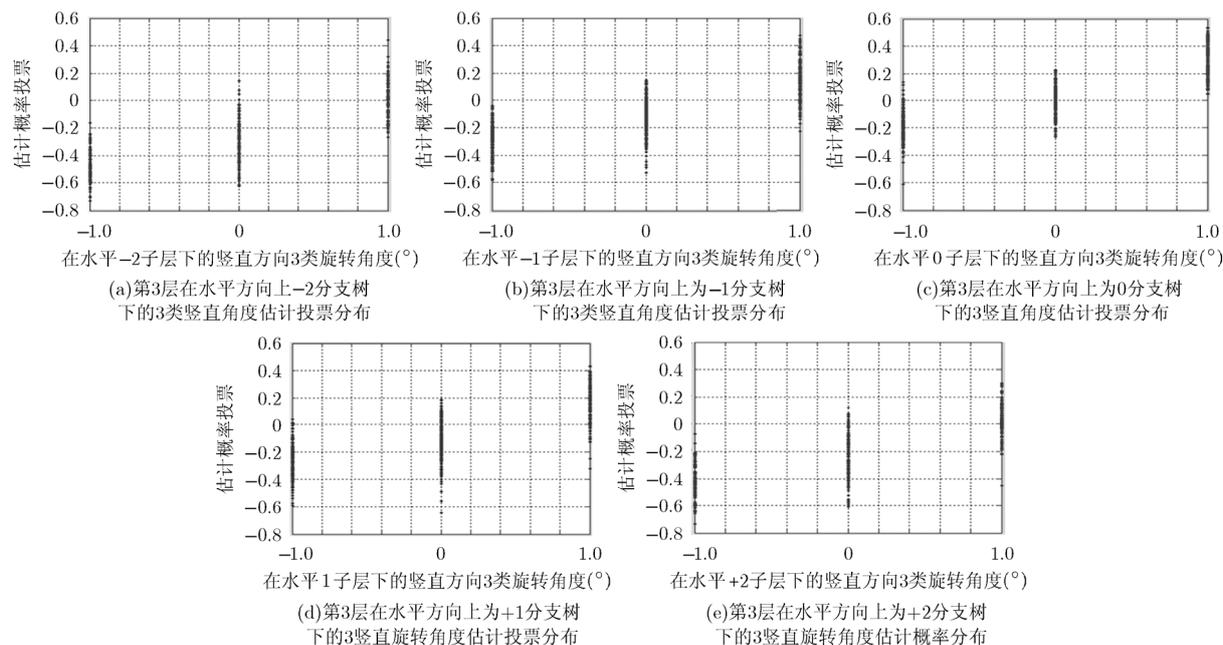


图 11 树结构分层随机森林的垂直自由度下的估计概率投票分布

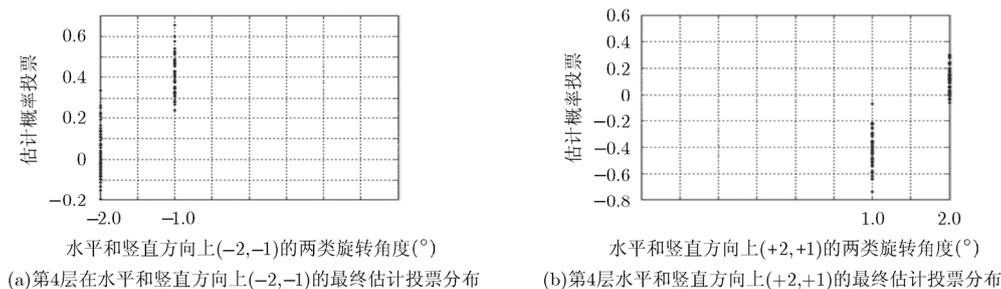


图 12 树结构分层随机森林的最终概率投票分布

表 1 随机森林(R)和本文算法(D)的估计准确率(%)

竖直 (°)	水平 (°)									
	90		45		0		-45		-90	
	D	R	D	R	D	R	D	R	D	R
90	72.1	61.0	69.7	62.5	72.3	71.6	68.3	55.6	70.6	65.9
45	73.0	52.3	72.6	73.1	73.2	69.4	71.9	43.9	71.5	66.2
0	79.3	75.2	75.9	64.1	78.7	70.6	74.0	69.8	80.0	75.7
-45	72.4	66.0	70.5	72.8	70.1	73.2	68.8	50.7	67.9	49.4
-90	67.2	58.8	70.3	60.3	70.7	67.0	69.4	45.2	65.3	60.4

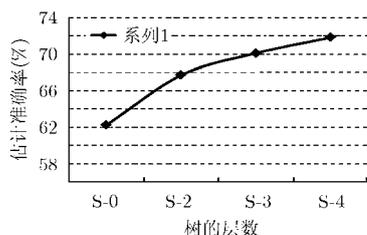


图 13 树结构分层随机森林的估计准确率与级联子树层的关系

67.72%, S-3 为增加 3 层树分布后得到的 25 类头部姿态估计准确率上升为 70.11%, S-4 为增加 4 层数分布的 25 类估计准确率提高到 71.38%。

4.5 估计准确率

表 2 给出了一些不同分类算法的比较实验结果。结果显示本文分层算法对两个自由度下的 25 类头部姿态估计有最佳的估计效果。

表 2 不同算法的准确率比较

不同算法	估计准确率(%)
本文分层算法	71.83
先验随机森林 ^[15]	70.11
条件随机森林 ^[6]	68.75
分类随机森林 ^[8]	64.23
SVM 分类算法 ^[16]	50.30

4.6 实时性比较

所有的实验都是基于硬件 PC Intel(R)Core (TM) i5-2400 CPU@ 3.10 GHz, 32 bit 的系统, 软件为 Microsoft Visual Studio2010++ 平台。随机选择数据库中的 200 张图片, 计算它们的平均运行时间。表 3 描述了本文算法和最新算法的平均运行时间 $\bar{\mu}$ 的比较。本文算法的运行时间最短。

表 3 本文算法与随机森林算法的运行时间比较 (s)

不同算法	前景-背景分类	水平头部姿态估计	竖直头部姿态估计	总运行时间
本文算法	0.206914	0.233516	0.352829	0.793259
条件随机森林	-	0.696799	0.402610	1.099409
分类随机森林	-	0.696799	0.671794	1.368594
SVM 分类算法	-	1.032584	0.802897	1.835481

5 结束语

本文提出了一种能提高鲁棒性和准确性的头部姿态估计方法, 对于非约束环境下的两个自由度旋转的多类头部姿态估计有较好的效果。首先, 为了消除不同环境的噪声影响, 我们提取人脸区域的组

合纹理特征, 对人脸区域进行积极人脸子区域的分类, 分类结果作为树结构分层随机森林的先验知识输入; 第二, 提出了一种树结构分层随机森林算法, 分层估计多自由度下的头部姿态; 第三, 为了增强算法的分类能力, 本文使用自适应高斯混合模型作为多层次子森林叶子节点的投票模型。最后 3 个不同数据集上的测试实验表明本文方法比现有的先进算法具有更好的估计准确率和实时性。下一步我们将针对大场景下的多人头部姿态估计, 建立估计模型, 对于头发和眼睛的遮挡建立更加鲁棒的特征模型。

参考文献

- [1] Zhang M, Li K, and Liu Y. Head pose estimation from low-resolution image with Hough forest[C]. 2010 IEEE Chinese Conference on Pattern Recognition (CCPR), Chongqing, China, 2010: 1-5.
- [2] Cai Q, Sankaranarayanan A, Zhang Q, *et al.* Real time head pose tracking from multiple cameras with a generic model[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), San Francisco, CA, 2010: 25-32.
- [3] Martinovi A, Mathias M, Weissenberg J, *et al.* A Three-layered Approach to Facade Parsing[M]. Computer Vision-ECCV 2012, Springer Berlin Heidelberg, Florence, Italy, 2012: 416-429.
- [4] Guo G, Fu Y, Dyer C R, *et al.* Head pose estimation: Classification or regression?[C]. 19th IEEE International Conference on Pattern Recognition, Tampa, FL, USA, 2008: 1-4.
- [5] Murphy-Chutorian E and Trivedi M M. Head pose estimation in computer vision: a survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(4): 607-626.
- [6] Dantone M, Gall J, Fanelli G, *et al.* Real-time facial feature detection using conditional regression forests[C]. 25th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island, 2012: 2578-2585.
- [7] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5-32.
- [8] Huang C, Ding X, and Fang C. Head pose estimation based on random forests for multiclass classification[C]. 20th IEEE International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 2010: 934-937.
- [9] Fanelli G, Gall J, and Van Gool L. Real time head pose estimation with random regression forests[C]. 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, 2011: 617-624.
- [10] Fanelli G, Weise T, Gall J, *et al.* Real Time Head Pose

- Estimation from Consumer Depth Cameras[M]. England, Pattern Recognition, Springer Berlin Heidelberg, 2011: 101-110.
- [11] Li Y, Wang S, and Ding X. Person-independent head pose estimation based on random forest regression[C]. 17th IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 2010: 1521-1524.
- [12] Minka T. The dirichlet-tree distribution[OL]. <http://www.stat.cmu.edu/minka/papers/dirichlet/minka-dirtree.pdf>, 1999.
- [13] Gourier N, Hall D, and Crowley J L. Estimating face orientation from robust detection of salient facial structures[C]. FG Net Workshop on Visual Observation of Deictic Gestures, Cambridge, UK: FGnet (IST-2000-26434), 2004: 1-9.
- [14] Huang G B, Mattar M, Berg T, *et al.* Labeled faces in the wild: a database for studying face recognition in unconstrained environments[C]. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, 2008: 1-14.
- [15] Yang H and Patras I. Privileged information-based conditional regression forest for facial feature detection[C]. 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 2013: 1-6.
- [16] Ba S O and Odobez J M. Multiperson visual focus of attention from head pose and meeting contextual cues[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(1): 101-116.
- 刘袁缘：女，1984年生，硕士，讲师，研究方向为计算机视觉与模式识别、图像处理。
- 陈靓影：女，1973年生，博士，教授，研究方向为计算机视觉与模式识别、图像处理。
- 俞侃：男，1978年生，博士，副教授，研究方向为光学图像处理。