

# 基于形状索引的 DoG 特征结合 GPRT 的人脸关键点检测算法

赵康

(商丘职业技术学院, 河南 商丘 476100)

**摘要:** 针对传统人脸对齐算法效率较低的问题, 提出一种基于形状索引的高斯差分 (DoG) 特征与高斯过程回归树 (GPRT) 的人脸关键点检测算法; 首先, 由高斯过程回归树的内核测量两个输入之间的相似性, 并表示为两个输入进入相同叶子的树木数; 然后基于高斯过程回归树模型提取形状索引 DoG 特征, 并进一步完成 GPRT 的特征设计; 最后从局部视网膜模式中采集滤波回来增加稳定性, 实现对抗几何差异的鲁棒性; 在 LFPW 人脸数据库上验证结果表明该方法能够取得良好的性能表现, 证明了基于形状索引的 DoG 特征与 GPRT 的人脸关键点检测算法的有效性。

**关键词:** 高斯模型; 人脸关键点检测; 形状索引; 鲁棒性; 几何差异

## Shape Index Based DoG Feature and GPRT Face Key Detection Algorithm

Zhao Kang

(Shangqiu Vocational and Technical College, Shangqiu 476100, China)

**Abstract:** Aiming at the low efficiency of traditional face alignment algorithm, a face index based Gaussian difference (DoG) feature and Gaussian process regression tree (GPRT) face key detection algorithm are proposed. First, the kernel between the two inputs is measured by the kernel of the Gaussian process regression tree and is expressed as the number of trees entering the same leaf for both inputs. Then, the shape index DoG feature is extracted based on the Gaussian process regression tree model, and the feature design of the GPRT is further completed. Finally, the filtering response is collected from the local retina mode to increase the stability and achieve robustness against geometric differences. The verification results on the LFPW face database show that the proposed method can achieve good performance, and proves the validity of the shape index-based DoG feature and GPRT face key detection algorithm.

**Keywords:** gaussian model; face key point detection; shape index; robustness; geometric difference

## 0 引言

人脸关键点检测就是对面部特征进行精确定位, 如面部图像中的眼角、鼻尖、嘴角和下巴。准确且鲁棒地人脸对齐有利于实现各种涉及人脸的目标应用, 如人脸识别<sup>[1-2]</sup>、面部表情识别<sup>[3-4]</sup>等。但是由于人脸形状本身具有多变性, 以及光照、姿态、遮挡等因素的影响, 准确且鲁棒的人脸关键点检测还具有相当的难度。

近年来, 级联形状回归模型在特征点定位上取得了重大突破。形状回归是一种人脸对齐框架<sup>[5]</sup>, 其具有准确性、鲁棒性及快速性等特点。在形状回归中, 脸部形状是一个连接的面部基准坐标, 其通过一连串分类与回归树 (Classification and Regression Tree, CRT) 实现原始迭代更新。每一个树基于当前的形状评估对形状增量进行估量, 最终

的形状评估由原始评估的输出树累计加和而来。目前, 已经在形状回归中考虑了不同种的正则化方法以便降低过度拟合, 获取更好的泛化性能。文献 [7] 通过对一个人脸图像生成多个初始形状估计来增强训练数据, 这一数据加强的方法已经在后续的研究中应用。文献 [8] 将收缩和平均作为正则化方法, 在梯度推进学习过程中, 学习率参数乘以每一个回归树或独立学习和平均树。文献 [9] 将学习过程分为两个步骤: (1) 学习二进制映射函数 (2) 学习现行回归矩阵。二进制映射函数包括一组局部的二进制映射函数, 该函数由单面部基准点的方法从独立学习树中诱导。接着通过平方损失最小化函数, 获取线性回归矩阵。然而这些算法易陷入局部极值并且鲁棒性不强。

本文使用高斯过程回归树 (Gaussian Process Regression Tree, GPRT) 来代替梯度推进。GPRT 通过一种特殊的内核简化预测过程中的计算, 在相同的预测时间内比 CRT 的泛化性能高。GPRT 的预测平均值为树输出的总和, 这一方法的预测时间相同但是提高了泛化性能。此处将 GPRT 的预测平均值设计成正比于预测变量 (来自于

收稿日期: 2018-11-05; 修回日期: 2018-11-28。

基金项目: 河南省高等学校重点科研项目(15A520118); 河南省科技厅软科学研究计划项目(142400411213)。

作者简介: 赵康(1983-), 男, 河南商丘人, 讲师, 硕士, 主要从事图像处理、算法设计等方向的研究。

GPRT 集合) 的值, 由此引出 GPRT 的贪婪分布式学习方法。在参考形状评估的局部视网膜模式中, 计算高斯函数的差分 (Difference of Gaussian, DoG) 特征来确定 GPRT 的输入特征。提取 DoG 特征, 首先, 在不同尺度条件下使用高斯滤波器平滑面部图像从而降低噪音敏感性。然后从光滑面部图像中提取像素值, 利用局部视网膜抽样模式、形状评估和光滑尺度把图像编入索引内。最后计算提取像素值的差异。每一个局部视网膜抽样点的光滑尺度与采样点和中心点之间的距离成正比。因此, 远距离采样点比附近的采样点涉及的区域面积大, 对形状评估误差的稳定性也较好, 识别能力也更强。

## 1 高斯过程回归树

### 1.1 回归树模型

CRT 中考虑了一组树, 并将形状回归看作成一种加法级联形式的树, 如下所示:

$$\hat{s}^T = \hat{s}^0 + \sum_{t=1}^T f^t(x'; \theta) \quad (1)$$

其中:  $t$  是代表阶段的指数,  $\hat{s}^t$  是形状评估,  $x'$  是从输入图像  $I$  中提取出来的特征向量,  $f^t(\cdot; \cdot)$  是  $\theta$  参数化的树。从粗略的原始形状评估  $\hat{s}^0$  开始, 每一个阶段通过  $\hat{s}^t = \hat{s}^{t-1} + f^t(x'; \theta)$  迭代更新形状评估。

考虑到训练样本  $S = (s_1, \dots, s_N)^T$  和  $X^t = (x'_1, \dots, x'_N)^T$ , 采用回归残差和贪婪分布式方法来学习树以便减小平方损失:

$$\theta = \underset{\theta}{\operatorname{argmin}} \sum_{i=1}^N \|r'_i - f^t(x'; \theta^*)\|_2^2 \quad (2)$$

此处, 回归残差为  $r'_i = s_i - \hat{s}_i^{t-1}$ 。

树参数  $\theta$  包括分割函数  $\tau^t(x')$  和回归输出  $\{\hat{r}^{t,b}\}_1^B$ 。分割函数的输入为  $x'$ , 计算叶指数  $b \in \{1, \dots, B\}$ , 每一个回归输出与相对应的叶指数  $b$  相关。对叶子上所有下降的训练数点进行平均回归残差得出最佳的回归输出:

$$\hat{r}^{t,b} = \frac{1}{N^{t,b}} \sum_{i: \tau^t(x')=b} r'_i \quad (3)$$

其中:  $N^{t,b}$  是叶子  $b$  上下降的训练数据点。用分割函数和回归函数改进的公式 (1) 为:

$$\hat{s}^t = \hat{s}^{t-1} + \hat{r}^{t, T(x')}$$

### 1.2 GPRT 的过程

GPRT 由许多 GPRT 组成, 每一个 GPRT 都有一个由树组定义的内核。在 GPRT 中, 输出和输入之间的关系由回归函数  $f(x)$  表示, 该回归函数是一个有独立加性噪音  $\epsilon_i$  的高斯过程:

$$s_i = f(x_i) + \epsilon_i, i = 1, \dots, N \quad (4)$$

$$f(x) \sim GP(0, k(x, x')) \quad (5)$$

$$\epsilon_i \sim N(0, \epsilon_n^2) \quad (6)$$

给定一个测试输入  $x$ , 在预测变量  $f_*$  上的分布为:

$$p(f_* | x_*, X, S) = N(f_* | \bar{f}_*, \sigma_*^2) \quad (7)$$

$$\bar{f}_* = k_*^T K_s^{-1} S \quad (8)$$

$$\sigma_*^2 = k_* - k_*^T K_s^{-1} k_* \quad (9)$$

其中:  $k_*$  和  $k_*$  分别是  $x_*$  和  $X_*$  之间的  $k(x_*, x_*)$  及协方差向量。此处,  $K_s$  由  $K + \sigma_n^2 I_N$  提供,  $K$  是协方差矩阵,  $K(i, j)$  由  $X$  的  $i, j$  排行向量计算得来。预测平均值也可以写成  $N$  个内核的线性组成:

$$\bar{f}_* = \sum_{i=1}^N \alpha_i k(x_i, x_*) \quad (10)$$

其中:  $\alpha = (\alpha_1, \dots, \alpha_N)^T$  由  $K_s^{-1} S$  提供。GPRT 的内核  $k(x, x')$  由含有  $M$  树的集合定义, 与文献 [8] 中的方法类似:

$$k(x, x') = \sigma_k^2 \sum_{m=1}^M k^m(x, x') \quad (11)$$

$$k^m(x, x) = \begin{cases} 1 & \text{if } \tau^m(x) = \tau^m(x') \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

其中:  $\sigma_k^2$  是代表内核能力的尺度参数, 该内核在树 (两个输入在树上落入相同的叶子) 数量的基础上计算两个输入之间的相似性。

GPRT 的超参数  $\sigma_k^2$  和  $\sigma_n^2$  由梯度优化的方法评估, 该优化方法对边缘相似性取对数。

$$\log p(S | X, \sigma_k^2, \sigma_n^2) = -\frac{1}{2} S^T K_s^{-1} S - \frac{1}{2} \log |K_s| - \frac{n}{2} \log 2\pi \quad (13)$$

在不损失泛化性能的前提下, 超参数  $\sigma_k^2$  和  $\sigma_r^2 = \frac{\sigma_n^2}{\sigma_k^2}$  可以替换  $\sigma_k^2$  和  $\sigma_n^2$ 。为了将边缘相似性的对数最大化来设置  $\sigma_k^2$ , 本文对  $\sigma_r$  取偏导数。

$$\frac{\partial}{\partial \sigma_r} \log p(S | X, \sigma_k^2, \sigma_r^2) = \frac{1}{2} \operatorname{tr}((\alpha \alpha^T - K_s^{-1}) \frac{\partial K_s}{\partial \sigma_r}) \quad (14)$$

公式 (13) 和 (14) 的计算负担包括计算  $K_s^{-1}$  和  $\log |K_s|$ , 这些都是函数  $O(N^3)$  所需的。然而, 逆运算却相当顺利, 因为  $K$  的排名是  $BM$  树上的最大叶子数。假设  $q_i = (q_i^1, \dots, q_i^M)^T, q_i^m$  是 *one-of-B* 代码向量, 其索引为第  $m$  树 (第  $i$  个训练点下降) 的叶节点。接着  $K = \sigma_k^2 Q Q^T$ , 其中  $Q = (q_1, \dots, q_N)^T$ 。由此可得:

$$K_s^{-1} = \sigma_k^{-2} (\sigma_r^{-2} I_N - \sigma_r^{-2} Q K_r^{-1} Q^T) \quad (15)$$

$$\log |K_s| = N \log \sigma_k^2 + (N - BM) \log \sigma_r^2 + \log |K_r| \quad (16)$$

其中的逆运算为  $O(BM^3), K_r = Q^T Q + \sigma_r^2 I_{BM}$  是  $BM \times BM$  矩阵。

当对  $\sigma_r^2$  评估后  $\sigma_k^2$  也可由类似的方式评估:

$$\sigma_k^2 = \frac{S^T (\sigma_r^{-2} I_N - \sigma_r^{-2} Q K_r^{-1} Q^T) S}{N} \quad (17)$$

在 GPRT 中, 输入量  $f_*$  的预测变量  $x_*$  是高斯随机变量, 其预测均值和变量分别由公式 (10) 和 (9) 提供。公式 (10) 的计算参考函数  $O(N)$ 。但是下述方法更高效:

$$\bar{f}_* = \sum_{i=1}^N \alpha_i k(x_i, x_*) =$$

$$\sum_{m=1}^M \sum_{i=1}^N \alpha_i \sigma_k^2 k^m(x_i, x_*) = \sum_{m=1}^M \bar{\alpha}^{m, \tau^m(x_i)} \quad (18)$$

其中： $\bar{\alpha}^{m,b} = \sigma_k^2 \sum_{i, \tau^m} \alpha_i$  是所有  $\alpha_i$  的总和，相对应的  $x_i$  落入叶子  $b$ 。更直观地说， $\bar{\alpha}^{m,b}$  可以解释为伪输入的预测平均值，伪输入只落在第  $m$  棵树的叶子  $b$  上。

此外，为了测量树上每一片叶子的预测不准确性，假设  $(\bar{\sigma}^{m,b})^2$  是上述伪输入的预测变量。

通过公式 (18)，在  $O(M \log B)$  中计算预测平均值，该值的计算框架与 CRT 中的预测相同。级联 GPRT 包含  $T$  个 GPRT，并结合 GPRT：

$$p(f_* | x_*, M) \propto \prod p(f_* | x'_*, M^t) \quad (19)$$

其中： $M$  和  $M^t$  是 GPRT 模型和  $T$  个 GPRT 模型。因为 GPRT 的每一个预测变量都是高斯随机变量，平均值为  $\{\bar{f}_*^t\}_{t=1}^T$ ，变量为  $\{(\sigma_*^t)^2\}_{t=1}^T$ 。GPRT 的预测随机变量  $f_*$  仍是一个高斯随机变量，平均值和变量分别定义为：

$$\bar{f}_* = \sigma_*^{-2} \sum_{t=1}^T (\sigma_*^t)^{-2} \bar{f}_*^t \quad (20)$$

$$\sigma_*^2 = \left( \sum_{t=1}^T (\sigma_*^t)^{-2} \right)^{-1} \quad (21)$$

在公式 (20) 中， $\bar{f}_*$  由  $\{\bar{f}_*^t\}_{t=1}^T$  的加权求和定义，权重代表了每一个 GPRT 预测的不确定性。

公式 (20) 中预测平均值的相加形式引发了 GPRT 的贪婪分布式学习方式，其中使用了回归残差。每一个 GPRT 进行再加权，再加权是为了建立当前剩余模型  $r^t$ ，其是回归函数  $f^t$  的求和，接下来的子序列残差  $r_i^{t+1}$  是高斯随机变量，平均值为零，变量为  $\sigma_n^2$ 。

## 2 形状索引的 DoG 特征的提取与学习

提取 DoG 特征的过程如下：

1) 对目标图像进行高斯过滤<sup>[11-12]</sup>，得到模糊的光滑图像如图 1 所示。

2) 计算相似性变换，将平均形状映射给形状评估。

3) 将相似性变换应用到局部视网膜采样模式，如图 2 所示。

4) 使用变换后的局部视网膜抽样模式并参考形状评估来计算全局坐标。

5) 在高斯光滑图像的全局坐标中提取像素值，从而对应于每一个采样点的尺度参数，提取高斯滤波器响应。

提取出来的两个高斯滤波器响应之间的差异体现在形状索引的 DoG 特征上，然后计算预定义的 DoG 过滤的响应。将相似性转换应用到局部视网膜采样模式，形状索引的 DoG 特征计算并不包含整个图像的转换，只转换稀疏的坐标<sup>[5]</sup>。另外，获取高斯光滑图像的计算复杂性并不高，因为光滑过程只有一次，且在学习过程之前。

用单一的面部坐标来学习 GPRT 树，由随机采样阈值和参考第 1 面部坐标的 DoG 特征获取树的分割函数。为了获取更有识别力的分割函数，测试了一些分割函数并选取



图 1 高斯模糊效果

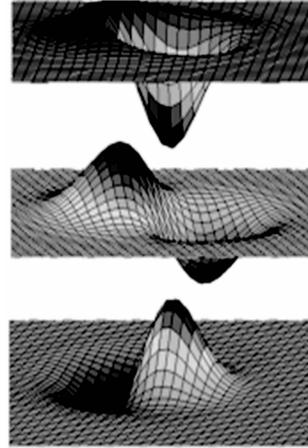


图 2 局部视网膜采集模式过程

了性能最佳者。而分割函数的性能由第 1 面部坐标的平方损失来衡量。这一过程仅学习了树的分割函数，而回归输出由 GPRT 学习。每一个树基于当前的形状评估对形状增量进行估量，最终的形状评估由原始评估的输出树累计加和而来。形状回归中影响预测性能的两个关键因素，分别是学习 CRT 的梯度推进和形状检索特征。从梯度推进得来的 GPRT 通常过度拟合。在梯度推进中，每一个阶段以贪婪分步实施方式，迭代性地拟合训练数据，该过程减少了回归残差，这些残差的定义由地面真实形状与形状评估之间的差异来完成。在学习和预测的过程中存在拟合率之间的差异时，便会出现过度拟合，在一些阶段中拟合训练数据的速度太快，可能会导致较差的泛化性能，在预测过程中也会出现不精准的形状评估。

最初阶段的学习树倾向于使用形状索引的 DoG 特征，这些特征由距离较远的采样点计算得来，而处于后期的树倾向于使用从附近的采样点计算得来的特征。这是因为较远的采样点设计更大的区域，对于形状评估误差来说更稳定。附近的采样点稳定性差一点，但是当形状评估准确时识别力高。因此，在学习阶段，形状索引的 DoG 特征允许每一个树根据实际需要自由选择更稳定的特征。

## 3 实验结果与分析

本实验的目的有：(1) 将使用了形状索引的 DoG 特征的 GPRT 与最先进的方法进行比较。(2) 证实本文方法的

关键因素：GPRT 和形状索引 DoG 特征。

### 3.1 LFPW 数据库

本文大部分实验结果都基于 LFPW 数据库<sup>[10]</sup>，它是知名的目前最先进最具挑战性的数据库。LFPW 数据库包含 1132 个训练图像和 300 个测试图像，其能提供链接到图像的 URL。本文收集了 778 个训练图像和 216 个测试图像。LFPW 人脸数据库部分截图如图 3 所示。



图 3 LFPW 人脸数据库部分截图

将形状评估误差看作是内目镜距离的一部分，该距离是经瞳孔距离标准化后的地面实际情况和形状评估间距离。所有的实验结果都选取平均值（超过 10 次的实验）以便降低随机性的影响。

### 3.2 实验设置

为了获取训练数据，首先使用面部检测边界框来裁剪人脸图像。接着，形状评估初始化成随机采样的真实形状，该形状来自其他训练数据点。在训练阶段，每一个面部图像的初始化过程要重复二十次。预测过程中使用来自训练数据点的平均形状进行初始化。

本文考虑两种构型配置：(1)“GPRT”配置，具有较低的平均误差但是预测也较慢；(2)“GPRTfast”配置，预测较快但是平均误差大。在 GPRT 配置中，每一个 GPRT 的树数量和 GPRT 的数量分别为  $M = 10$  和  $T = 500$ 。GPRT 由两种层级的级联 GPRT 组成，第一层级的阶段数和第二层级的阶段数分别为 100 和 5。在 GPRTfast 配置中，每个 GPRT 的树数量和 GPRT 的数量分别设置为  $M = 10$ ,  $T = 100$ 。第一层级的级联阶段数和第二层级的级联阶段数都设为 10。对两个配置而言，树的深度设为 5。每一个分割函数都要经过 200 次试验，光滑尺度设为 8。每一个光滑尺度中视网膜采样点数为 6，因此每一个面部坐标的采样点数为  $6 \times 7 + 1 = 43$ 。所以试验的操作系统为：配备单核 i5-3570、3.40 Hz CPU 的 PC 机。

### 3.3 与先进算法之间的比较

在这一节，将使用了形状索引的 DoG 特征的 GPRT 与以下最先进的方法进行比较：形状回归 (ESR)<sup>[13]</sup>，鲁棒的级联构成回归<sup>[14]</sup>、碱度下降法 (SDM)<sup>[15-16]</sup>。其中文献 [13] 提出一种形状到形状的显式形状回归方法，利用丰富的几何形状先验定位人脸特征点。并且引入一种射影不变量—特征数 (CN) 去描述和刻画人脸固有的几何形状结构。

提出一种鲁棒、有效的特征点检测器。通过构建特征数与形状之间的映射关系，建立从局部到全局的形状回归模型。LFPW 人脸关键点部分检测结果如图 4 所示。文献 [14] 重点对鲁棒关键点定位技术进行了对比研究，并设计实现了一个综合鲁棒表情关键点定位演示系统。文献 [15] 提出了基于人眼拓扑结构的人眼特征点跟踪。在 LFPW 人脸数据集进行试验，对比几种方法，结果如表 1 所示。



图 4 LFPW 人脸关键点检测

表 1 LFPE 数据集上不同方法的比较

方法	准确率(100%)
形状回归(ESR)	77.64
鲁棒的级联构成回归	74.46
碱度下降法(SDM)	68.68
本文方法	84.93

从表 1 可以看出，本文方法取得了 84.93% 的准确率，而其他几种方法均低于 80%。LFPW 数据库的实验结果表明 GPRT 的性能明显优于其他方法。当使用形状检索特征时，过度拟合会更重要。形状检索特征与形状评估紧密相关：形状检索特征决定了形状评估，而形状检索特征又是从基于图像评估的像素坐标中提取出来。拟合率之间的差异导致预测过程中提取出来的形状检索特征都不相关，进而导致提取出的特征更不相关。其中文献 [13] 提出的基于形状回归 (ESR) 的方法以及文献 [15] 提出的基于碱度下降法 (SDM) 的方法均依赖于建立模型，对于存在遮挡或者模糊状态下的面部图像往往难以建立准确的模型，制约了方法的使用。而文献 [14] 提出的基于鲁棒的级联构成回归的方法，在定位鲁棒关键点时，易受到算法本身计算成本过高的干扰，对于复杂场景中的面部图像准确率还有待进一步提升。

## 4 结论

本文提出了一种基于形状索引的 DoG 特征与 GPRT 相结合的人脸关键点检测算法。GPRT 由 GPRT 集合组成，并以一种贪婪分布式方法学习。GPRT 的预测均值可在 CRT 框架中计算，泛化性能更佳。而且，本文也描述了形状索引 DoG 特征。参考形状评估，在局部视网膜模式上计算高斯滤波响应差异，并由此来设计上述的形状索引 DoG 特征。使用该特征的 GPRT 在 LFPW 数据库获得最佳性能。为了验证本文方法的有效性，将其与现有的几种方法进

(下转第 211 页)