

快速人脸检测与特征定位

朱文佳 戚飞虎

(上海交通大学计算机科学与工程系, 上海 200030)

摘要 人脸检测与特征定位是人脸分析技术的一个重要组成部分,其目标是在图像中搜索人脸特征(如眼、鼻、嘴、耳等)的位置。虽然人们可以毫不费力地完成这些工作,但对于机器来说,这依然是一件极其困难的任务。近几年来该项技术已有了长足的发展,已成功地应用于诸如人脸识别、姿态识别、表情识别、脸部动画等诸多领域。本文利用 David Cristinace 和 Tim Cootes 提出的一个多阶段人脸特征检测方法实现了一个实时人脸特征定位系统。同时也对原算法本身作了一些改进,在对精度影响极小的情况下,大大提高了原算法的速度。

关键词 人脸检测 人脸特征定位

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2005)11-1454-04

Fast Face Detection and Facial Features Localization

ZHU Wen-jia, QI Fei-hu

(Dept. of Computer Science and Engineering, JiaoTong University, Shanghai 200030)

Abstract Face detection and facial features localization are important parts of face analysis techniques. It's task is to search facial features in an image. Although human can finish this work without any effort, it's still a very difficult work for computer. In the past years, this technique has developed vastly and has been successfully applied in the fields of face recognition, pose recognition, expression recognition and face cartoon, etc. In this paper, we use a multi-stage facial features detection algorithm proposed by David Cristinace and Tim Coote to implement a real-time facial features localization system. Meanwhile, we also improve the origin facial features localization algorithm to run much faster with only a little loss in accuracy.

Keywords face detection, facial feature localization

1 引言

因为人脸的特征一方面存在着较大的个体间的差异,另一方面由于受光照,表情,遮挡等等因素的影响,所以人脸特征定位是一项十分困难的工作。举例来说,戴眼镜和不戴眼镜有很大差别,张大嘴和闭上嘴又有很大差别。

依目前技术水平而言,由于无法像人脸检测那样,在保持高检测率的同时,维持极低的误报率,所以需要利用一个方法来从多个响应点中筛选出正确的特征点。本文采用并改进了由 Cristinace 和

Cootes 提出的一个多阶段人脸特征检测算法^[1],较好地解决了这个问题。

2 方法概述

Cristinace 和 Cootes 于 2004 年提出了一个多阶段的人脸特征检测算法^[1]。该算法在一个公开的人脸库——BIOID 上获得了目前人脸特征定位的最高精度。该算法分为以下 3 个阶段。第 1 阶段是人脸检测,其目的为了是在图片中定位人脸的位置,以便人脸特征只需要在一个相对较小的区域内搜索,而不必在整张图片内搜索,这样就可以大大提高精

收稿日期:2005-08-20; 改回日期:2005-09-15

第一作者简介:朱文佳(1983 ~)男。2005 年获上海交通大学学士学位,现为上海交通大学计算机科学与技术系在读硕士研究生。主要研究方向为计算机视觉与模式识别。E-mail: zhuwenjia@ yahoo. com. cn

度和速度;第 2 阶段是人脸特征定位,即在已检测出的人脸区域中定位人脸各个特征的位置,这些特征主要包括:眉毛、眼睛、鼻子、嘴巴;第 3 阶段是用改良过的 AAM (active appearance model) 算法来进一步优化结果。为了达到实时性的目的,本系统仅采用了前两阶段。

第 1 阶段的人脸检测请参考文献[2]。本文的主要工作是对第 2 阶段的 PRFR (pairwise reinforcement of feature response) 算法的改进。

3 人脸特征定位

PRFR 算法,即特征响应的成对加强。该算法的基本思想是利用人脸特征的几何位置关系来实现在不同的情况下正确的人脸特征定位。人脸特征的位置虽然有较大的变化,但它们之间的相对位置总是满足一定的关系,比如眼睛一定在鼻子上面,内眼角必定比眼球更靠近中线等等。PRFR 算法就是利用了这个相对位置关系,使位置正确的点之间满足这种关系,互相确信;而错误的点不满足这个关系,因而无法互相确信。这样便可定位出正确的点。

具体地说,这个算法的实现可分为以下 2 个步骤:首先需要通过一个分类器来检测出各个特征点的可能位置,采用的仍然是与人脸检测阶段同样的分类器;然后通过统计得出的特征之间的相对分布为

$$P_{i,j}(x_i | x_j)$$

上式的含义是当第 j 个特征的响应点位于 x_j 时,第 i 个特征处于 x_i 的概率,但由于这样一个概率分布图实在过于庞大(由于 x_i, x_j 都表示一个 2 维点,因此 $P_{i,j}$ 是四变量的函数),所以实际使用中可用两变量的函数 $M_{i,j}(x_i - x_j)$ 代替相对分布图。 $M_{i,j}$ 的一些实例见图 1。

图中打叉的点是 $M_{i,j}$ 的 $(0,0)$ 点。在分布图中越是黑的点表示概率越大。

由图 1 可见,对于 $i=j$,也就是该特征的检测位置 and 实际位置和分布关系,这样的 $M_{i,j}$,其特征分布主要集中在 $(0,0)$ 左右,分布比较紧凑(见图 1(a),图 1(b)),这表明使用级联分类器来检测人脸特征基本上还是比较精确的。对于 $i \neq j$ 的情况,当这两个特征的相对位置比较稳定时, $M_{i,j}$ 的特征分布就比较紧凑,如外眼角相对眼球(见图 1(c));反之就比较分散,如眼球相对左嘴角(见图 1(d))。

$M_{i,j}$ 的训练过程如下所述:

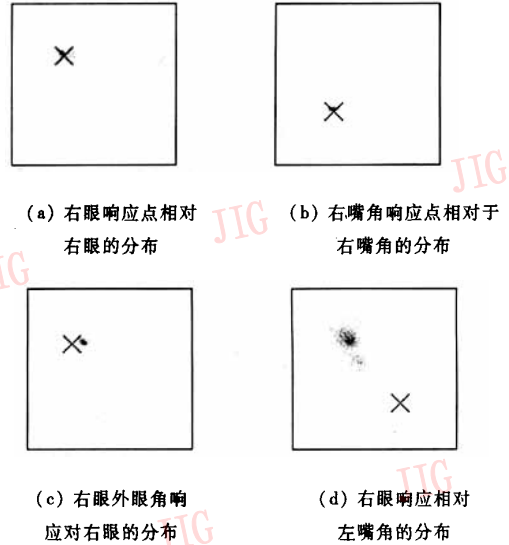


图 1 相对分布图 $M_{i,j}$ 的一些实例

Fig. 1 Some examples of $M_{i,j}$

(The point with a cross is $(0,0)$ of $M_{i,j}$.)

A blacker point indicates a greater value in $M_{i,j}$.)

- (1) 令 N 表示训练样本数量;
- (2) 置所有的 $M_{i,j}(d) = 0$;
- (3) For $k = 1 \rightarrow N$.

在第 k 张训练图片上,先使用各个特征检测器来提取最强响应点;然后记录这个点和其他所有的特征的真实点在位置上的相对偏差。假定偏差为 d ,则更新相应的 $M_{i,j}(d)$,即

$$M_{i,j}(d) = M_{i,j}(d) + 1/N \quad (1)$$

在训练阶段,当对训练样本进行统计确定了相对分布图 $M_{i,j}$ 后,下一步就是通过投票来确定特征点的正确位置。对于每个特征 i ,由下式来确定它的位置

$$x_i = \arg \max \sum_j \sum_t M_{i,j}(x_i - q_{j,t}) \quad (2)$$

式中, $q_{j,t}$ 表示第 j 个特征检测器的第 t 个最大响应。

这个式子的含义是,从每个响应点看来,特征 i 具有一定的概率分布,而 x_i 的正确位置就是从所有响应点看来概率最大的那个位置。举例来说,对于眼睛这个特征,眼睛检测器认为眼睛就在眼睛响应点附近,而从鼻尖检测器看来,眼睛应该在鼻尖响应点的上方,这样综合考虑了所有特征的响应点的“意见”之后,概率最大的那个点就被认为是眼睛。

具体实现的时候,只要根据 $M_{i,j}$ 和 $q_{j,t}$,就可以得到从 $q_{j,t}$ 看来,特征 i 的概率分布,然后将所有

这样的分布累加起来,再选取可能性最大的点,就得到了特征 i 的位置。

该算法的优点在于巧妙地利用了人脸特征点的位置关系,从而有效地筛选出了正确的点。举例来说,常常发生的情况是嘴角检测器误把鼻孔附近认为是嘴角,然而经过投票之后,由于其他特征的正确响应点绝不会认为嘴角是在鼻孔附近,因此,误报的嘴角响应点的确信度较低,而正确的嘴角响应点的确信度则较高,这样正确的嘴角就被选出来了。

4 PRFR 算法改进

虽然 PRFR 算法巧妙地使用了人脸特征的形状约束,其不仅能有效的从多个候选点中筛选出正确的特征点,并且具有很强的健壮性;然而该算法的缺点也很明显,假如每个特征点取 3 个最强的响应点,那么在投票的时候,每确定一个特征就需要累加 17×3 张分布图(取 17 个特征点)。这样总的计算量为 $17 \times 17 \times 3 = 867$ 张分布图的累加, Cristinacce 在论文中^[1]指出,这一步最为耗费时间,其在 500MHz 的 PC 上需耗费整整 800ms。由于本文的目的是进行实时检测,所以必须对这一步加以改进。

笔者认为,改进的方向大致有以下两个。

(1)在确定一个特征时,是否需要所有特征响应点一起投票?如果只用一半特征来投票,则速度就可以快一倍,然而究竟依照何种标准来决定哪些点参与投票而那些点又不参与投票呢?由于对整个测试集测试一遍需要几分钟的时间,所以通用的优化算法(如遗传算法)等似乎也不行。

(2)降低 $M_{i,j}$ 的精度。Cristinacce 将整个人脸划分为 100×100 的区域来建立 $M_{i,j}$,精度较高。假如降低一些精度,如取 50×50 ,那么速度一下子就可以快 4 倍。如果再降低一些精度,如取 35×35 ,那么速度又可以快近 9 倍。

通过以上分析不难看出,第 2 个思路既简单易行,又能有效减小运算量,所以最终本文采用了第 2 种思路。

改良后的算法借用了图像处理领域中非常经典的从粗到精(coarse-to-fine)的策略来建立了双精度的 $M_{i,j}$ 。首先用 35×35 的精度划分人脸,建立一个精度较低的概率分布图 $M_{i,j}^{\text{low}}$ 。再用 100×100 的精度划分人脸,建立一个精度较高的概率分布图 $M_{i,j}^{\text{high}}$ 。

在投票的时候先用低精度的分布图 $M_{i,j}^{\text{low}}$ 投票,确定各个特征点的大致位置,再用高精度分布图 $M_{i,j}^{\text{high}}$ 在其周围小块区域内(3×3)投票,以进一步优化结果。这样一来,既保持了原算法的健壮性与精确性,又把速度提高了近 9 倍。该算法的示意图见图 2,其实际效果将会在实验部分说明。



(a)在粗精度 $M_{i,j}^{\text{low}}$ 上
确定眼睛的大致位置

(b)在高精度 $M_{i,j}^{\text{high}}$ 上
确定眼睛的最终位置

图 2 双精度 PRFR 算法示意图

Fig. 2 Double-precision PRFR

5 实验结果

本实验采用了人脸库 AR^[23]中每个人的前 5 张图片,共 680 张图片作为特征检测器训练样本。为达到更好的训练效果,每个局部特征经随机旋转 15° 以内,采样 4 次。这样每个特征共有 2720 个训练样本。而测试样本则采用 BIODID 人脸数据库。

本文通过将检测到的特征点与手动标注的特征点进行比较来评估算法的精度。特征点的平均误差计算如下:

$$\bar{e} = 1/(ns) \times \sum_{i=1}^n d_i \quad (3)$$

上式中, d_i 是第 i 个特征检测点到真实点的误差; s 是两眼之间的距离; n 是特征点的数量。在本实验对每个特征取前 3 个最强的相应点(这是 Cristinacce 的推荐参数,可在健壮性和进度之间获得一个较好的平衡)。原始的 PRFR 算法和改进的 PRFR 算法的精度评估的结果如图 3 所示。

从上面的数据可见,改进的 PRFR 算法在精度上与原算法差别不大。只在误差小于 10% 的部分稍差。举例来说,对于误差为 10% 时,原算法的成功率为 75.5%,改进算法的成功率为 74.2%,虽然成功率有微小的差异,但改进后算法的耗费时间却大大减少,仅为原算法的 1/9。

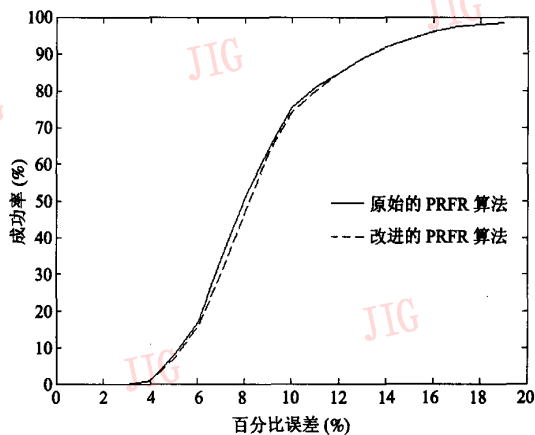


图 3 原始 PRFR 和改进的 PRFR 算法成功率比较

Fig. 3 Comparison between the origin PRFR and improved PRFR

6 结论

本文应用 Cristinace 和 Cootes 提出的一个多阶

段的人脸特征检测方法构建了一个实时的人脸特征定位系统并对原方法中最为耗时的 PRFR 算法进行了研究与改进。原算法的优点是利用了人脸特征的相对位置信息,因此能有效地从多个响应点中筛选出正确点,排除误报点,但其计算量过大,耗时过久。为此本文利用从粗到精的策略,提出了双精度的 PRFR 算法,改进后的新算法可在基本不影响精度的前提下,大大提高算法的速度,使之能够实时运行。

参考文献 (References)

- 1 Cristinace David, Cootes Tim, Scott Ian. A multi-stage approach to facial feature detection [A]. In: British Machine Vision Conference [C], Kingston University UK, London 2004, 1:277 ~ 286.
- 2 Viola Paul, Jones Michael. Robust real-time face detection [J]. International Journal of Computer Vision 2004, 57(2):137 ~ 154.
- 3 Martinez A M, Benavente R. The AR face database [R]. In: Computer Vision Center at U. A. B Technical Report #24. 1998 [EB/OL]. <http://rvll.ecn.purdue.edu/~aleix/ar.html>