

doi: 10.14098/j.cn35-1288/z.2018.02.008

基于深度值和 M-估计的人脸姿态估计

邱丽梅^{1,2}, 邱思杰¹

(1. 三明学院 机电工程学院, 福建 三明, 365004;

2. 机械现代设计制造技术福建省高校工程研究中心, 福建 三明, 365004)

摘要:对于单幅人脸图像姿态估计, 针对现有的估计方法大多存在缺乏人脸特征深度信息而造成姿态估计过程病态化的缺陷, 及最小二乘法不能很好地区分内外特征点, 从而影响估计精度和鲁棒性的问题。提出利用透视原理和迭代算法计算人脸特征点的深度值, 在最小二乘法的基础上, 加入 M-估计优化算法, 精确估计出人脸三维空间姿态。实验结果表明, 该方法不仅克服了姿态估计的病态化问题, 而且具有良好的姿态估计精度和鲁棒性。

关键词:人脸姿态估计; 深度值; 特征点; 最小二乘法; M-估计算法

中图分类号: TP391.41

文献标志码: A

文章编号: 1673-4343(2018)02-0049-07

Face Pose Estimation Based on Depth and M- estimation

QIU Li-mei^{1,2}, QIU Si-jie¹

(1. School of Mechanical & Electronic Engineering, Sanming University, Sanming 365004, China;

2. Fujian University Engineering Research Center of Modern Mechanical Design and Manufacturing Technology, Sanming 365004, China)

Abstract: Most of the existing pose estimation methods have the defect that the process of pose estimation is ill from lack of the depth information, and cannot distinguish the internal and external points of the characteristics because of the least square method, affecting the accuracy and robustness of estimation. Therefore, a new approach of using the principle of perspective and iterative algorithm to calculate the depth of facial feature points is proposed in this paper for estimating 3D face pose from a single image. Then, the M-estimation algorithm is added on the basis of the least square method and the 3D face pose is estimated accurately. Experimental results show that this method not only overcomes the ill posed problem of pose estimation, but also has good accuracy and robustness of pose estimation.

Key words: face pose estimation; depth value; feature points; least-squares; M- estimation algorithm

近年来, 虽然随着深度摄像机技术的发展及微软 Kinect 的问世, 打破了以往深度摄像头高昂价格的局面, 人脸姿态估计领域也出现了 ETH Face Pose 和 Biwi Kinect 两个新的深度图像数据库^[1], 但在海关、机场、展览馆等公共场合及公安追捕罪犯系统等实际应用中, 对于单幅二维人脸图像的多姿态识别仍居主流^[2]。目前, 正面人脸识别系统已取得了较好的识别效果, 但是多姿态人脸图像的有效识别, 效果仍然欠佳。据统计大约 75% 的人脸图像为非正面的人脸图像^[2], 除此以外, 人脸姿态还在智慧城市、司机疲劳检测系统和智能轮椅等方面都有着广泛的应用^[3]。

根据实现原理和方式的不同, 现有的人脸姿态估计方法大体上可以分为六类^[4]: (1) 形状模板匹配的方法^[5-6]; (2) 检测分类器的方法^[7-8]; (3) 特征回归的方法^[9-10]; (4) 流形嵌入的方法^[11-12]; (5) 局部约束

收稿日期: 2017-12-08

基金项目: 福建省自然科学基金计划项目(2017J01778); 福建省科技项目(JK2015045); 三明市科技项目(2016-G-8);

福建省科技厅科技项目(2014H0004)

作者简介: 邱丽梅, 女, 福建清流人, 副教授。主要研究方向: 模式识别、机电一体化。

模型的方法^[13]; (6)脸部特征点几何关系的方法^[14-16]。其中,脸部特征点几何关系的方法具有简单、耗时短、效率高等优点,而且随着人脸特征点提取算法的日趋成熟,脸部关键特征点的检测和定位精度已逐渐获得保障^[17]。

针对单幅人脸图像,依据以下几点:(1)多姿态二维人脸图像识别仍居主流;(2)脸部特征点几何关系方法具有简单、耗时短、效率高等优点;(3)现有的姿态估计方法大多存在缺乏人脸特征深度信息而造成姿态估计过程病态化的问题;(4)最小二乘法不能很好地区分内点和外点,提出了基于深度值和 M-估计的人脸姿态估计方法。

1 求解人脸特征点的深度值

对于单幅人脸图像在不知道人脸特征点深度值及其它信息的前提下,只利用脸部特征点几何关系估计人脸空间姿态是不易实现的。若能得知人脸特征点的深度值,就能丰富人脸预知信息,精准实现对单张照片中人脸的姿态估计。

采用 CMU PIE 人脸图像为研究对象,进行了相关预处理后,得到大小为 64×64 的图像。先利用改进的活动形状模型(Active Shape Model,简称 ASM)方法提取人脸特征点^[17],再利用透视原理和迭代算法计算得到人脸特征点的深度值。

假设一特征点的 3D 空间坐标为 $y_n=(y_{n1} \ y_{n2} \ y_{n3})^T$ 和相应的 2D 投影坐标为 $u_n=(u_{n1} \ u_{n2})^T$, 旋转矩阵 $R=(r_1 \ r_2 \ r_3)^T$, 平移矢量 $t=(t_1 \ t_2 \ t_3)^T$, f 为像机焦距,则:

$$u_{n1} = f \frac{r_1 y_n + t_1}{r_3 y_n + t_3}, \quad u_{n2} = f \frac{r_2 y_n + t_2}{r_3 y_n + t_3} \quad (1)$$

在相机坐标系中,根据透视成像有:

$$u_n = \begin{pmatrix} u_{n1} \\ u_{n2} \\ f \end{pmatrix} = f \begin{pmatrix} v_{n1} \\ v_{n2} \\ 1 \end{pmatrix} = f v_n \quad (2)$$

其中 $v_{n1} = \frac{r_1 y_n + t_1}{r_3 y_n + t_3}, v_{n2} = \frac{r_2 y_n + t_2}{r_3 y_n + t_3}$ 。

y_n 经过旋转 R , 平移 t 后得到 x_n , 则:

$$x_n = \begin{pmatrix} x_{n1} \\ x_{n2} \\ x_{n3} \end{pmatrix} = R \begin{pmatrix} y_{n1} \\ y_{n2} \\ y_{n3} \end{pmatrix} + t \quad (3)$$

在相机坐标系中,假设 x_n 的深度值为 $d_n, n=1,2,3, \Lambda, N$ 则:

(1) 选取 x_n 的初始深度值为 d_n^0, d_n^0 是人脸特征点与透视中心的距离;

(2) 假设经过迭代,得到第 k 步的深度值 d_n^k , 那么第 $k+1$ 步的深度值 d_n^{k+1} , 可通过使得式(4)最小化求得:

$$\varepsilon_k^2 = \sum_{n=1}^N w_n \|R_k y_n + t_k - d_n^k v_n\|^2 \quad (4)$$

其中, ε^2 为方差, $\{w_n | n=1, \Lambda, N\}$ 是非零因子, 满足 $w_n > 0$, 在此定义 w_n 相等且 $w_n = 1/N$ 。 R_k 和 t_k 分别是第 k 步时所对应的人脸旋转矩阵和平移矩阵, 即人脸姿态值。

(3) 假设式(4)收敛, 即有: $\varepsilon_{k+1}^2 \leq \varepsilon_k^2$, 证明如下:

$$\varepsilon_{k+1}^2 = \sum_{n=1}^N w_n \|R_{k+1} y_n + t_{k+1} - d_n^{k+1} v_n\|^2$$

$$\begin{aligned}
&\leq \sum_{n=1}^N w_n \|R_k y_n + t_k - d_n^{k+1} v_n\|^2 \\
&= \sum_{n=1}^N w_n \|(x_n^k - d_n^k v_n) + (d_n^k v_n - d_n^{k+1} v_n)\|^2 \\
&= \sum_{n=1}^N w_n [\|(x_n^k - d_n^k v_n)\|^2 + 2(x_n^k - d_n^k v_n)^T \cdot (d_n^k - d_n^{k+1}) v_n + (d_n^k - d_n^{k+1})^2 \|v_n\|^2] \\
&= \varepsilon_k^2 + \sum_{n=1}^N w_n (d_n^k - d_n^{k+1}) \cdot [2(x_n^k - d_n^k v_n)^T v_n + (d_n^k - d_n^{k+1}) \|v_n\|^2] \\
&= \varepsilon_k^2 + \sum_{n=1}^N w_n (d_n^k - d_n^{k+1}) \cdot [2x_n^{kT} - 2d_n^k \|v_n\|^2 + (d_n^k - d_n^{k+1}) \|v_n\|^2] \\
&= \varepsilon_k^2 + \sum_{n=1}^N w_n \|v_n\|^2 \left[(d_n^{k+1})^2 - 2 \frac{x_n^k v_n^T}{\|v_n\|^2} d_n^{k+1} + 2 \frac{x_n^k v_n^T}{\|v_n\|^2} d_n^k - (d_n^k)^2 \right] \\
&= \varepsilon_k^2 + \sum_{n=1}^N w_n \|v_n\|^2 \left[\left(d_n^{k+1} - \frac{x_n^k v_n^T}{\|v_n\|^2} \right)^2 - \left(d_n^k - \frac{x_n^k v_n^T}{\|v_n\|^2} \right)^2 \right]
\end{aligned} \tag{5}$$

可见,要使式(5)取得最小值,那么:

$$d_n^{k+1} = \frac{x_n^k v_n^T}{\|v_n\|^2} \tag{6}$$

此时,(5)式就变成了:

$$\varepsilon_{k+1}^2 \leq \varepsilon_k^2 - \sum_{n=1}^N w_n \|v_n\|^2 \left(d_n^k - \frac{x_n^k v_n^T}{\|v_n\|^2} \right)^2 \tag{7}$$

显然,由式(7)可得, $\varepsilon_{k+1}^2 \leq \varepsilon_k^2$ 成立,所以:

$$d_n^{k+1} = \frac{x_n^k v_n^T}{\|v_n\|^2} \tag{8}$$

把式(3)代入(8)式,可得:

$$d_n^{k+1} = \frac{x_n^k v_n^T}{\|v_n\|^2} = \frac{(R_k y_n + t_k)^T v_n}{v_n^T v_n} \tag{9}$$

2 基于深度值和 M-估计的人脸姿态估计

由于特征点提取误差不可避免的会产生个别误差较大的外点,最小二乘法不能很好地区分内点和外点,从而影响估计精度和鲁棒性,因此,在最小二乘法的基础上,加入 M-估计算法进一步优化。

随机的估计值 T_k ,可由下式最小化形式表示:

$$\min_{T_k} \sum_{i=1}^n \rho(x_i - T_k) \tag{10}$$

或可通过下列隐式方程表示:

$$\sum_{i=0}^n \Psi(x_i - T_k) = 0 \tag{11}$$

其中, ρ 是原函数:

$$\Psi(x-T_k) = \frac{\partial}{\partial T_k}(x-T_k) \quad (12)$$

把式(12)写成以下等式形式:

$$\sum_{i=0}^n w_i(x_i-T_k)=0 \quad (13)$$

其中:

$$w_i = \frac{\Psi(x_i-T_k)}{x_i-T_k}, \quad i=1,2,3,\Lambda,N \quad (14)$$

所以, T_k 可以被表示成:

$$T_k = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (15)$$

本文采用 Huber 的 ρ 和 Ψ 算子,表示如下:

$$\rho(x) = \begin{cases} 0.5x^2, & |x| \leq a \\ a|x| - 0.5a^2, & |x| > a \end{cases} \quad (16)$$

$$\Psi(x) = \begin{cases} -a, & x < -a \\ x, & |x| \leq a \\ a, & x > a \end{cases} \quad (17)$$

a 是调整常数,取残差绝对值的常数倍,在此处 $a=1.5$ 。

非线性回归问题可以描述如下:

假设函数 $f_i: E^m \rightarrow E, i=1,2,3,\Lambda,n; \theta=(\theta_1, \theta_2, \Lambda, \theta_m)^T \in E^m$ 是未知的 m 维向量,则:

$$f_i(\theta) = y_i, i=1,2,3,\Lambda,n \quad (18)$$

求解方程式(18),即求解式(19)

$$\theta = \min \sum_{i=1}^n \rho\left(\frac{y_i - f_i(\theta)}{S}\right) \quad (19)$$

有数种不同的方法,为了构造一种呈比例变化的 M-估计,估计比例因子取值如下:

$$S = \frac{\text{median}_i |y_i - f_i(\theta)|}{0.6745} \quad (20)$$

通过修正残差值的方法来解式(19)最小问题,确定 θ 的步骤如下:

- (1) 选择初始的 θ^0 ;
- (2) 假设第 k 步的迭代值是 θ^k ,则按以下步骤来解决:
 - (a) 计算残差值 $r_i^*, i=1,2,3,\Lambda,n$,则:

$$r_i^* = \Psi\left(\frac{r_i}{S^k}\right) S^k$$

其中: $r_i = y_i - f_i(\theta^k), S^k = \text{median}_{r_i \neq 0} |r_i| / 0.6745$ 。

- (b) 求解最小二乘问题 $X\delta = r^*$, 其中, $X = \begin{vmatrix} x_{ij} \end{vmatrix}$ 是梯度矩阵,即:

$$x_{ij} = -\frac{\partial}{\partial \theta_j} f(\theta^k)$$

采用一般的最小二乘法便可求解此方程;如果采用奇异值分解法,则 $X=U \sum V^T$,那么其解为 $\delta = V \sum^{-1} U^T r^*$;

(c) $\theta^{k+1} = \theta^k + \delta$ 。

3 实验

3.1 实验步骤

本次实验在 Matlab 2012a 环境下编程所得,其实现步骤如图 1 所示。

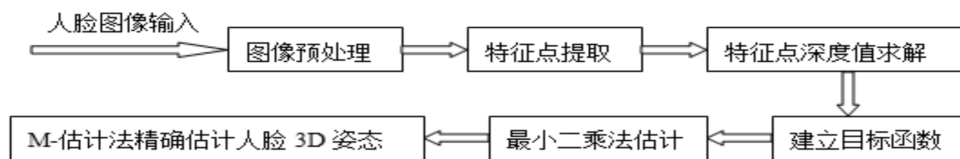


图 1 人脸姿态估计流程图

- (1) 输入待估计姿态人脸图像。
- (2) 把输入的彩色图像转化成灰度图像、灰度值归一化为[0,1],并进行去噪和标准化处理,图像大小统一处理为 64×64。
- (3) 对处理好的人脸图像,引入仿射变换不变性初始化原则(ATIIA)建立初始模型,改进传统的活动形状模型(ASM)方法^[17],提取左右眼内外角点、左右嘴角点、鼻尖点、两鼻翼点和下巴顶点等共十个主要人脸特征点的二维值。
- (4) 根据透视成像原理,利用迭代算法,求得特征点的深度值。
- (5) 选用左右眼内外角点、左右嘴角点、鼻尖点、两鼻翼点和下巴顶点等共十个主要特征点来表示人脸,利用透视成像原理,建立人脸姿态估计目标函数。
- (6) 利用两外眼点、鼻下点和嘴角点等四个特征点进行姿态一次估计,作为对目标函数进行最小二乘优化运算的输入值,输出二次估计的姿态结果,并把它们作为下步 M-估计的初始值。
- (7) 利用 M-估计算法对最小二乘法得到的估计值作进一步优化,输出最终估计结果。

3.2 实验数据

实验采用 CMU PIE 人脸数据库,该数据库包含了 68 个人的脸,均为彩色图像,以 PPM 格式存储,分辨率为 640×486,人脸图像大小不一,背景多样,4 种不同表情,43 种不同光照,13 个离散姿态,共 41368 幅人脸图像。

实验前,我们把彩色图像转化成灰度图像、灰度值归一化为[0,1],并进行去噪和标准化处理,图像大小统一处理为 64×64。对其中的 $\beta=0^\circ, \pm 22.5^\circ, \pm 45^\circ$ 5 个姿态共 170 幅人脸图像进行了测试,其中也包含了部分“自遮挡”(眼镜和少量头发遮挡)。随机选取了部分实验图像,如图 2 所示。

3.3 实验结果与分析

一般的应用系统中,对绕 Y 轴旋转(β)的侧面图像应用较多,因此,随机选取部分真值为 $\beta=0^\circ, \pm 22.5^\circ, \pm 45^\circ$ 这 5 种姿态图像,如图 3 所示。图中第一组数据为本文算法(最小二乘法+M-估计优化算法)估计得出的姿态值,第二组数据为最小二乘算法估计得出的姿态值,结果表示形式为 (α, β, γ) ,单位为度($^\circ$),随机选取部分测试结果如图 3 所示。



图 2 部分预处理后的实验图像

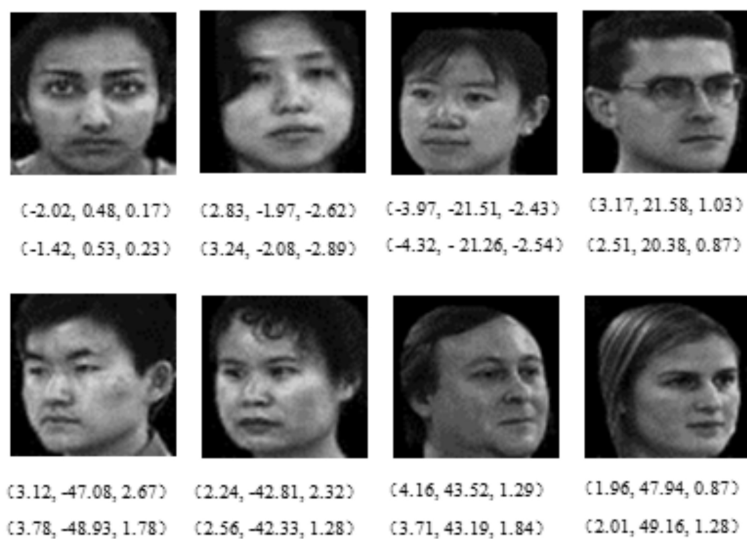


图 3 部分估计结果

对于人脸姿态估计的性能评价指标通常有:平均绝对误差、绝对误差标准差和方向估计精度 3 种^[4],平均绝对误差是估计精度最为简单的评价指标,也是估计系统通用的指标,一般误差在 5° 以内,认为是有效估计。表 1 中分别给出了本文算法和最小二乘法的 β 平均绝对误差。

表 1 β 的平均绝对误差对比表

算法	平均绝对误差/($^\circ$)				
	0	+22.5	-22.5	+45	-45
本文算法	1.18	2.28	2.31	2.76	2.94
最小二乘法	1.29	2.32	2.42	3.98	4.22

通过对比图3和表1可以发现:两种算法的姿态估计误差随着角度的增加而增大,其中真值为 $\beta=0^\circ, \pm 22.5^\circ$ 时,两种算法估计误差相差不大,都控制在 3° 内的高精度估计;真值为 $\beta=\pm 45^\circ$ 的情况下,本文算法误差仍稳定地控制在 3° 内的高精度,而最小二乘算法则出现了不稳定现象,误差时而在 3° 以内,时而超过 4° ,但平均绝对误差也都控制在 5° 以内的有效估计。

4 结论

本文提出的基于深度值的单幅图像人脸姿态估计方法,利用了透视原理和迭代算法计算出特征点的深度值,克服了现有姿态估计方法大多存在缺乏人脸特征深度信息而造成姿态估计过程病态化的缺陷;利用M-估计算法解决了最小二乘法不能很好地区分内点和外点的问题,保证了估计精度和鲁棒性。通过CMU PIE人脸数据库进行实验,结果表明:对于 $\beta=-45^\circ \sim +45^\circ$ 的单幅二维人脸图像,本算法具有良好的估计精度和鲁棒性。

参考文献:

- [1] JOHN DARBY, MARI' A B SA'NCHEZ, PENELOPE B. BUTLER, et al. An evaluation of 3D head pose estimation using the Microsoft Kinect V2[J]. Gait Posture, 2016, 48(4): 83-88.
- [2] 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[J]. 软件学报, 2009, 20(6): 1651-1663.
- [3] WU SIYU, LIANG JIE, JASON HO. Head pose estimation and its application in TV viewers' behavior analysis[C]//Electrical and Computer Engineering. IEEE, 2016: 1-6.
- [4] 唐云祁, 孙哲南, 谭铁牛. 头部姿势估计研究综述[J]. 模式识别与人工智能, 2014, 27(3): 213-225.
- [5] PADELERIS P, ZABULIS X, ARGYROS A A. Head pose estimation on depth data based on particle swarm optimization [C]//Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. 2012: 42-49.
- [6] NHAT QUAN, HUYNH NGUYEN, THAI HOANG LE. Pose estimation using local binary patterns for face recognition[C]//Springer International Publishing Switzerland 2015.
- [7] ROBERTO VALLE, JOS'E MIGUEL BUENAPOSADA, ANTONIO VALDE'S, et al. Head-pose estimation in the-wild using a random forest[C]//Springer International Publishing Switzerland 2016.
- [8] SRJJA CHOWDHURY, JAYA SIL. Head pose estimation for recognizing face images using collaborative representation based classification[C]//Proc of the Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016: 571-575.
- [9] FANELLI G, DANTONE M, GALL J, et al. Random forests for real time 3D face analysis[J]. International Journal of Computer Vision, 2013, 101(3): 437-458.
- [10] MA B P, CHAI X J, WANG T J. A novel feature descriptor based on biologically inspired feature for head pose estimation[J]. Neuro-computing, 2013, 115(4): 1-10;
- [11] LU J W, TAN Y P. Ordinary preserving manifold analysis for human age and head pose estimation[J]. IEEE Trans on Human Machine Systems, 2013, 43(2): 249-258.
- [12] 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[J]. 中国图象图形学报, 2012, 17(8): 1002-1010.
- [13] ZHU X X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild[C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2879-2886;
- [14] HARI C V, PRAVEEN SANKARAN. A clustered locally linear approach on face manifolds for pose Estimation[C]//Springer Verlag London 2016.
- [15] LIANG X N, TONG W Q. Face pose estimation using near-infrared images[C]//Proc of the International Conference on Communication Systems and Network Technologies, 2012: 216-220.
- [16] ANWAR SAEED, AYOUB AL HAMADI, SEBASTIAN HANDRICH. Advancement in the head pose estimation via depth-based face spotting[C]//IEEE Symposium Series on Computational Intelligence, 2016.
- [17] WAN K W, LAM K M, CHONG K. An accurate active shape model for facial feature extraction[J]. Pattern Recognition Letters, 2005, 26(12): 2409-2423.

(责任编辑:朱联九)