

基于多变量标签分布的连续型姿态估计方法

江 婷,赵启军,陈 虎

(四川大学 计算机学院,四川 成都 610065)

摘 要:人脸的头部姿态往往指示并传达着丰富的信息,准确估计头部姿态角度在人脸识别、表情识别等领域有重要作用。针对获得的人脸真实姿态角度往往存在一定的偏差且只包含有限个离散角度等问题,文中提出了一种基于多变量标签分布的连续型姿态估计方法。在训练阶段,对不同姿态角度,通过训练获得离散情况下的多变量标签分布;在测试阶段,采用正交多项式拟合的思想,将离散的分布拟合成连续的分布,计算分布的最大值所对应的标签作为最终的输出结果。文中在 Pointing'04 公开库上进行了测试,利用文中方法,在正交多项式拟合后,估计出的人脸的姿态角不再局限于训练集中的一些角度,而是有更多连续的值,所得的估计姿态角更接近于真实角度。实验结果表明,文中方法能够预测出更多的人脸姿态角度,并且预测更稳定。

关键词:姿态估计;多变量标签分布;正交多项式拟合;连续型姿态

中图分类号:TP399.1

文献标识码:A

文章编号:1673-629X(2016)01-0111-04

doi:10.3969/j.issn.1673-629X.2016.01.023

A Continuous Head Pose Estimation Method Based on Multivariate Label Distribution

JIANG Ting,ZHAO Qi-jun,CHEN Hu

(School of Computer Science,Sichuan University,Chengdu 610065,China)

Abstract:The human's head pose are abundant of information. Accurate estimation of head pose plays an important role in face recognition,expression recognition and so on. To improve the precision of estimation and to alleviate the problem that poses are always fixed to some angles,a continuous method based on multivariate label distribution to estimate head poses was presented. In the training phase,get the discrete multivariate distribution from discrete poses and angles. In the testing phase,adopt orthogonal polynomial fitting to transform the discrete distribution into continuous distribution and compute the label corresponding to maximum in distribution as final output. The proposed method has been tested on the open Pointing'04 database. After orthogonal polynomial fitting,the estimated angles are no longer limited to angles in test set,but more continuous values. The estimated angle obtained by the method proposed is closer to the real angle. The result indicates that this method can estimate head pose in wider angle,the result is more stable.

Key words:head pose estimation;multivariate label distribution;orthogonal polynomial fitting;continuous head pose

0 引 言

在人际交往中,头部姿态是非常重要的部分,静态的头部姿态的某个特定方向可指示对某人的注意,或者对话的对象,等等。头部姿态的变化也可传达丰富的信息,比如赞同、否定、理解、疑惑和惊喜等等。另外,头部姿态是很多应用领域的关键部分,如人脸识别、表情识别、视线估计等等。因此,头部姿态估计已成为计算机视觉和模式识别的一个重要研究领域。近年来已经有很多头部姿态估计的方法被提出,例如非

线性回归方法^[1-5]、子空间嵌入方法^[6-9]、基于特殊特征的方法^[10-13]和多变量标签分布方法^[14]。

在计算机视觉领域,头部姿态估计通常指的是使用图像平面去预测头部方向。通常做法是假设人的头部是一个刚性物体,这样头部姿态只有三个自由度,包括水平偏转角、俯仰角和旋转角^[15]。对于很多现存的数据集,只考虑了有限的离散角度,而且姿态角度标定值往往存在一定的误差。文献[14]提出离散的多变量标签分布方法(Multivariate Label Distribution,

收稿日期:2015-04-29

修回日期:2015-08-04

网络出版时间:2016-01-04

基金项目:国家自然科学基金资助项目(61202160,61202161);科技部重大仪器专项(2013YQ49087904)

作者简介:江 婷(1991-),女,硕士研究生,研究方向为模式识别、计算机视觉;赵启军,副教授,硕士生导师,研究方向为模式识别、机器学习、计算机视觉等;陈 虎,讲师,硕士生导师,研究方向为模式识别。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160104.1505.026.html>

MLD), 将人脸姿态角度的临近角度也作为姿态估计的标签, 使得一个人脸姿态可以有不同权重的多个值, 提高了算法对训练数据中姿态角度标定误差的鲁棒性, 也提高了姿态估计的正确率。但是现有人脸数据库中的姿态角度往往是离散的值, 若估计到的姿态角是真实值的临近角度, 也会偏差很大。换言之, 现有的 MLD 方法对训练集中不包含的姿态角的图片预测误差往往会很大。

为解决上述问题, 文中提出了基于 MLD 的连续值姿态估计方法。首先用 MLD 得到离散情况下的多变量标签分布, 然后用正交多项式拟合的方法拟合成连续的分布。选用正交多项式作为拟合工具可以得到与一般多项式拟合相同的结果, 而且有效避免了一般多项式拟合时法方程组的病态问题。改进后的 MLD 估计得到的姿态角度比离散情况更接近真实值, 而不受训练集角度间隔的影响。值得一提的是, 提出的方法能非常方便地扩展到三个自由度或者缩减到单个自由度。

1 多变量标签分布

设一个姿态角 y (y 是一个由水平偏转角和俯仰角构成的二维向量) 相对于一张人脸图像 x 的描述度用 d_x^y 表示。假设 x 的真实姿态角为 \hat{y} , $d_x^{\hat{y}}$ 是所有可能姿态中的最大值。由于临近角度间表情的相似性, y 的临近姿态 \hat{y} 应该有一个非零且低于 $d_x^{\hat{y}}$ 的描述度 d_x^y , 并且描述度是随着 y 到 \hat{y} 的距离增加而减小的。假设 $d_x^y \in [0, 1]$ 且 $\sum_y d_x^y = 1$, 则对于一张图片, 其所有可能的姿态的描述度构成 MLD。MLD 允许用多个姿态角来表示一张图片, 这种方法对于那些不能精确定义其姿态角的图片更符合现实。

为了生成一张给出的图片 x 合理的 MLD, 假设 MLD 服从离散的、二维的以姿态 \hat{y} 为中心的高斯分布, 如:

$$d_x^y = \frac{1}{2\pi \sqrt{|\Sigma|} Z} \exp\left(-\frac{1}{2} (y - \hat{y})^T \Sigma^{-1} (y - \hat{y})\right) \quad (1)$$

其中, Z 为确保 $\sum_y d_x^y = 1$ 的归一化因子; Σ 为 2×2 的协方差矩阵, $\Sigma = \begin{bmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{bmatrix}$, 其中 τ 为姿态角的最佳间隔。

用上述方法生成 MLD 后, 训练集就成为 $G = \{(x_1, P_1), (x_2, P_2), \dots, (x_n, P_n)\}$ 。其中, P_i 为 x_i 的 MLD。假设有 n_p 个不同的俯仰角和 n_y 个水平偏转角,

则第 j 个俯仰角和第 k 个水平偏转角所对应的姿态用 y_{jk} 表示, 则 P_i 可用一个第 j 行 k 列值为 $d_{x_i}^{y_{jk}}$ (即 y_{jk} 相对于 x_i 的描述度) 的 $n_p \times n_y$ 的矩阵表示。

设在连续空间 $X = \mathbb{R}^q$ 中, 标签集 $Y = \{y_{jk}; j = 1, 2, \dots, n_p, k = 1, 2, \dots, n_y\}$, 那么目标就是从 G 学习得到一个条件函数集 $p(y|x; \theta)$ 。其中 $x \in X, y \in Y, \theta$ 为参数向量。该条件函数集由找到一个 θ 使生成的 MLD 接近于当前 x_i 的 P_i 来决定。由于 $p(y|x; \theta)$ 的形式与 Geng 等在文献 [16-17] 中的工作类似, 所以假设它为一个最大熵模型, 即:

$$p(y_{lm}|x_i; \theta) = \frac{1}{\Gamma_i} \exp\left(\sum_r \theta_{lmr} x_i^r\right) \quad (2)$$

其中, $\Gamma_i = \sum_{l,m} \exp\left(\sum_r \theta_{lmr} x_i^r\right)$ 为归一化因子; x_i^r 表示 x_i 的第 r 个特征; θ_{lmr} 为与姿态 y_{lm} 和第 r 个特征相关的集合 θ 中的一个元素。

2 正交多项式拟合

针对不同姿态角度, 训练获得离散情况下的多变量标签分布后, 采用正交多项式拟合的思想, 将离散的分布拟合成连续的分布, 计算分布的最大值所对应的标签作为最终姿态的估计值。

2.1 一元函数拟合

给定一系列观测点 $(x_i, y_i), i = 0, 1, \dots, n-1$, 假设正交多项式 $f(x) = \sum_{k=0}^{n-1} \mu_k \omega_k(x)$ 是拟合这些点的近似函数, 其中基函数为:

$$\begin{cases} \omega_0(x) = 1 \\ \omega_1(x) = x - \alpha_0 \\ \vdots \\ \omega_{k+1}(x) = (x - \alpha_k) \omega_k(x) - \beta_k \omega_{k-1}(x) \end{cases} \quad (3)$$

若令 $d_k = \sum_{i=0}^{n-1} \omega_k^2(x_i)$, 则

$$\begin{cases} \alpha_k = \sum_{i=0}^{n-1} x_i \omega_k^2(x_i) / d_k \\ \beta_k = d_k / d_{k-1} \\ \mu_k = \sum_{i=0}^{n-1} y_i \omega_k^2(x_i) / d_k \end{cases} \quad (4)$$

将式(4)代入式(3)中, 可求得 $\omega_k(x)$, 最后得到拟合多项式 $f(x) = \sum_{k=0}^{n-1} \mu_k \omega_k(x)$ 。

2.2 二元函数拟合

对于给定点 $(l_i, m_j) \rightarrow \theta_{ij}$, 其中 $i = 0, 1, \dots, L-1, j = 0, 1, \dots, M-1$ 。首先固定一个变量, 利用一元函数拟合对另一个变量进行拟合。首先固定 m , 对 m_0, m_1, \dots, m_{M-1} 进行 M 次关于 l 的拟合, 对 m_j 相应的拟合多

项式为 $g_j(l) = \sum_{k=0}^{p-1} \lambda_{kj} \omega_k(x)$ 。对 $g_j(l)$ 中每个固定 k 进行关于 (g_j, λ_{kj}) 的拟合, $j=0, 1, \dots, M-1$, 得到 p 个拟合的多项式: $h_k(m) = \sum_{s=0}^{q-1} \mu_{ks} \varphi_s(m)$, $k=0, 1, \dots, p-1$ 。所以 $\theta_{ij} = f(l_i, m_j) = \sum_{k=0}^{p-1} \lambda_{kj} \omega_k(l_i)$, 又因为 $\lambda_{kj} \approx h_k(m)$, 所以拟合的近似函数形式为:

$$f_{pq}(l, m) = \sum_{k=0}^{p-1} \sum_{s=0}^{q-1} \mu_{ks} \omega_k(l) \varphi_s(m) \quad (5)$$

先对 l 进行拟合后, 可得到 $d_k, \omega_k(l), k=0, 1, \dots, L-1$ 。再对 m 进行拟合, 得到 $\sigma_s, \varphi_s(m), m=0, 1, \dots, M-1$, 参考式(4), 则

$$\mu_{ks} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{M-1} \theta_{ij} \omega_k(l_i) \varphi_s(m_j)}{d_k \sigma_s} \quad (6)$$

将式(6)代入式(5)中即得拟合的二元函数。

当从一组可供选择的模型中选择一个最佳模型时, 选择赤池信息量准则 (Akaike Information Criterion, AIC) 为最小的模型是可取的, 所以采用 AIC 作为决定拟合多项式变量最高阶次数的方法。

$$AIC = N \times \ln \sigma^2(p, q) + 2pq \quad (7)$$

其中, $N = L \times M$; $\sigma^2(p, q)$ 为拟合多项式的极大似然函数。

拟合后, 在求得最优解 θ^* 后, 给出一张图片 x' , 则先根据 $p(y | x'; \theta^*)$, $y \in Y$ 估计得到它的 MLD, 然后与 MLD 中最大描述度相对应的姿态角即为 x' 的估计值。

3 算法流程

为了解决现有的 MLD 方法对训练集中不包含的姿态角的图片预测误差很大的问题, 提出了一种基于 MLD 的连续值姿态估计方法。首先在训练阶段, 对不

同姿态角度的图片进行训练, 将每个训练图片的临近姿态角度的权重设为非零值, 即该图片就能有多个标签值, 就可根据条件函数获得离散情况下的多变量标签分布。

在测试阶段, 采用正交多项式拟合的思想, 对于给定的测试图片, 通过将参数向量 θ 拟合成连续的函数, 获得 θ 的最优解, 然后与 MLD 中最大描述度相对应的姿态角即为该测试图片的估计值。整体框图如图 1 所示。

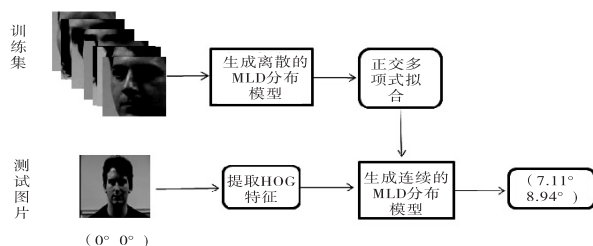


图1 基于 MLD 的连续型姿态估计方法的框图

4 实验结果及分析

为验证改进的 MLD 方法的估计效果, 将该方法在 Pointing'04 人脸库上进行了测试。Pointing'04 人脸库包括 9 个离散的俯仰角 $\{-90^\circ, -60^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ, 60^\circ, 90^\circ\}$, 以及 13 个离散的水平偏转角度 $\{-90^\circ, -75^\circ, -60^\circ, -45^\circ, -30^\circ, -15^\circ, 0^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ, 90^\circ\}$ 。人脸的姿态由一个水平偏转角和一个俯仰角组成, 特别地, 当俯仰角为 -90° 和 90° 时, 水平偏转角始终是 0° 。所以, 该数据库中包含 $13 \times 7 + 2 = 93$ 个姿态角。该库包含两次采集 15 个人的 人脸图像, 共 $93 \times 15 \times 2 = 2790$ 张。图 2 给出了该库中单个人的部分角度图像, 库中每张图片被归一化为 32×32 大小的灰度图, 用方向梯度直方图 (Histogram of Oriented Gradients, HOG) ^[18] 提取特征, 每个细胞单元为 3×3 个像素。

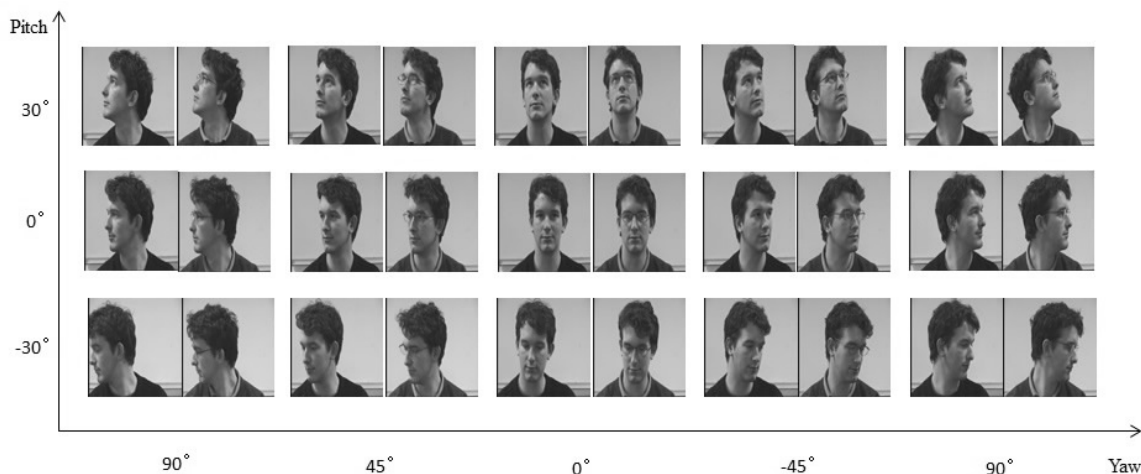


图2 Pointing'04 人脸库中同一个人部分角度图像

实验中,首先对 Pointing'04 人脸库随机选取 14 个人($93 \times 14 \times 2 = 2\,604$ 张)的图片进行训练,剩下 1 人($93 \times 2 = 186$ 张)图片进行测试。因为当俯仰角为 -90° 和 90° 时,水平偏转角始终是 0° ,所以测试时去除这两个姿态角,则剩余 91 个姿态。多次实验证明,当 $p = 4$, $q = 8$ 时,拟合得到的多项式最大地逼近了离散标签值。图 3 给出了对于测试图片,通过多项式拟合,阶数分别为 4 和 8 时, θ 前后值的对比,空心圆圈曲线表示 θ 的初始值,实心点曲线表示拟合后的 θ 的值,横轴表示 91 个姿态标签值,纵轴为 θ 的值。



图 3 θ 拟合前后曲线对比

在正交多项式拟合后,估计出的人脸的姿态角不再局限于训练集中的一些角度,而是有更多连续的值,表 1 列出了部分真实值所对应的两种方法估计出来的结果。

表 1 离散和连续 MLD 方法的估计结果对比

真实值		离散 MLD 估计值		连续 MLD 估计值	
60	75	30	60	60	79.121 54
60	0	90	0	60	-4.936 13
30	45	15	30	25.737 67	46.937 92
30	-60	15	-90	30.074 53	-60.252 7
15	60	0	75	11.699 75	74.923 76
15	-60	30	-75	19.786 8	-58.008 1
0	30	-15	15	0.753 241	36.939 08
0	0	-15	-15	7.115 74	8.938 457
-15	0	-30	-15	-14.030 7	2.116 103
-30	0	-30	-15	-32.834 9	9.210 472
-30	-60	-15	-60	-36.294 9	-64.647
-60	45	-60	30	-60	45.408 69
-60	-15	-60	-45	-50.903 5	-5.026 9

通过对比可以看出,连续的 MLD 方法估计出的值更接近真实值,而离散的方法使结果偏差为 15° 的倍数。

5 结束语

准确估计人脸的姿态角度在人脸识别、表情识别等领域有重要作用。文中受文献[14]的启发,针对受现有离散的人脸数据库中的有限姿态角度的影响,MLD 估计得到的姿态角也会偏差很大这个问题,提出了一种正交多项式拟合的方法,将标签拟合为连续值,使姿态估计地更接近真实值。实验结果表明,文中方法能够预测出更多角度的人脸姿态角度并且估计的姿态更稳定。

参考文献:

[1] Stiefelhagen R. Estimating head pose with neural networks – results on the Pointing04 ICPR workshop evaluation data [C]//Proc of pointing 2004 workshop: visual observation of deictic gestures. Cambridge, UK: [s. n.], 2004.

[2] Gourier N, Maisonnasse J, Hall D, et al. Head pose estimation on low resolution images [C]//Proc of CLEAR. [s. l.]: [s. n.], 2006: 270–280.

[3] Voit M, Nickel K, Stiefelhagen R. Neural network based head pose estimation and multi-view fusion [C]//Proc of CLEAR. [s. l.]: [s. n.], 2006: 291–298.

[4] Guo G, Fu Y, Dyer C R, et al. Head pose estimation: classification or regression [C]//Proc of 19th international conf on pattern recognition. FL: [s. n.], 2008.

[5] Haj M A, Gonzalez J, Davis L S. On partial least squares in head pose estimation: how to simultaneously deal with misalignment [C]//Proc of IEEE conf on computer vision and pattern recognition. [s. l.]: IEEE, 2012: 2602–2609.

[6] Tu J, Fu Y, Hu Y, et al. Evaluation of head pose estimation for studio data [C]//Proc of 1st international workshop on classification of events, activities and relationships. Southampton, UK: [s. n.], 2006: 281–290.

[7] Li Z, Fu Y, Yuan J, et al. Query driven localized linear discriminant models for head pose estimation [C]//Proc of IEEE international conf on multimedia and expo. Beijing, China: IEEE, 2007: 1810–1813.

[8] Foytik J, Asari V K. A two-layer framework for piecewise linear manifold-based head pose estimation [J]. International Journal of Computer Vision, 2013, 101 (2): 270–287.

[9] Lu Jiwen, 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.

[10] Gurbuz S, Oztop E, Inoue N. Model free head pose estimation using stereovision [J]. Pattern Recognition, 2012, 45 (1): 33–42.

[11] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild [C]//Proc of IEEE conf on computer vision and pattern recognition. [s. l.]: IEEE, 2012: 2879–2886.

司提供的 2 万条客户业务数据对模型进行验证。采取的方法是:随机抽取 1 000、2 000、5 000、10 000、20 000 条客户业务数据来预测潜在高流量客户,并将预测结果与实际结果进行比较,得到预测值与实际值的比值,从而验证模型的准确性。结果如图 4 所示。

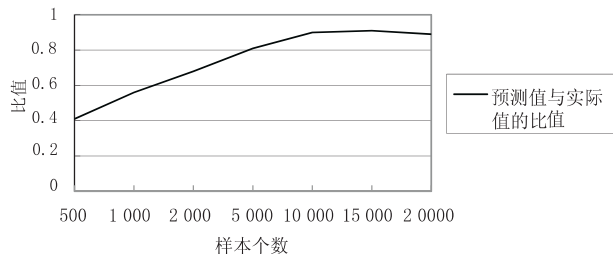


图 4 潜在高流量用户的预测检验

从图 4 中可以看出,当样本数 $<1\,000$ 时,由于偶然性大,预测值与实际值的比值小于 0.5,说明预测效果并不好。当样本数在 1 000 至 10 000 时,预测值与实际值的比值越来越大,并逐渐接近于 1,说明预测效果越来越好。当样本数在 10 000 到 20 000 之间时,预测值与实际值的比值趋于稳定并最接近于 1,说明预测效果最好。但是当样本数大于 20 000 后,模型的效果有略微下降趋势。

综上,样本数在 10 000 至 20 000 之间时,模型的预测效果较好,从而验证了模型的准确性。

5 结束语

文中通过对移动客户业务数据的预处理包括客户群的细分,建立了手机上网用户的决策树模型,并通过大量的测试数据对模型进行验证与评估,最后发现样本数据在 10 000 到 20 000 之间时预测效果较好。这说明该方法对于分类与预测潜在的高流量用户有较大的改进,从而能更好地为移动运营商适时推荐套餐,实现精确营销提供决策支持。但由于该样本数据集包含客户基本特征有限,例如客户学历、职业等特征的缺

少,文中所研究的内容还有待更进一步的深入。


参考文献:

- [1] 严霄凤,张德馨. 大数据研究[J]. 计算机技术与发展, 2013,23(4):168-172.
- [2] 范明,孟小峰. 数据挖掘概念与技术[M]. 北京:机械工业出版社,2001.
- [3] 陈志竞,梁伯瀚. 数据挖掘助力精细化流量经营[J]. 电信科学,2012,28(7):1-5.
- [4] 徐鹏,林森. 基于 C4.5 决策树的流量分类方法[J]. 软件学报,2009,20(10):2692-2704.
- [5] Han Hui, Mao Feng, Wang Wenyuan. Review of recent development in decision tree algorithm in data mining[J]. Application Research of Computers, 2004, 21(12):5-8.
- [6] 黄潇聪. 手机上网零流量用户“破零”模型的研究与应用[J]. 电信科学,2013(S2):26-29.
- [7] 董艳. 数据预处理方法在移动通信行业中的应用[J]. 计算机技术与发展,2010,20(11):225-228.
- [8] Fayyad U M. Data mining and knowledge discovery: making sense out of data[J]. IEEE Expert - Intelligent Systems & Their Applications, 1996, 11(5):20-25.
- [9] 张靖. 面向高维小样本数据的分类特征选择算法研究[D]. 合肥:合肥工业大学,2014.
- [10] 郭志懋,周傲英. 数据质量和数据清洗研究综述[J]. 软件学报,2002,13(11):2076-2082.
- [11] 赵蕊. 基于 WEKA 平台的决策树算法设计与实现[D]. 长沙:中南大学,2007.
- [12] Moore A W, Zuev D. Internet traffic classification using Bayesian analysis techniques[C]//Proc of SIGMETRICS. Banff: ACM, 2005:50-60.
- [13] Moore A W, Papagiannaki K. Toward the accurate identification of network applications[C]//Proc of LNCS. Heidelberg: Springer-Verlag, 2005:41-54.
- [14] 李霞. ID3 分类算法在银行客户流失中的应用研究[J]. 计算机技术与发展,2009,19(3):158-160.
- [15] Murphy-Chutorian E, Trivedi M M. Head pose estimation in computer vision: a survey[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2009, 31(4):607-626.
- [16] Geng X, Smith-Miles K, Zhou Z H. Facial age estimation by learning from label distributions[C]//Proc of 24th AAAI conf on artificial intelligence. Atlanta: [s. n.], 2010:451-456.
- [17] Geng X, Yin C, Zhou Z H. Facial age estimation by learning from label distributions[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(10):2401-2412.
- [18] Felzenszwalb P F, Girshick R B, McAllester D A, et al. Object detection with discriminatively trained part-based models[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2010, 32(9):1627-1645.

(上接第 114 页)

- [12] 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.
- [13] Ma B, Chai X, Wang T. A novel feature descriptor based on biologically inspired feature for head pose estimation[J]. Neuro-computing, 2013, 115:1-10.
- [14] Geng X, Xia Y. Head pose estimation based on multivariate label distribution[C]//Proc of IEEE conf on computer vision and pattern recognition. Columbus, Ohio: IEEE, 2014:1837-1842.
- [15] Murphy-Chutorian E, Trivedi M M. Head pose estimation in

基于多变量标签分布的连续型姿态估计方法

作者：[江婷](#)，[赵启军](#)，[陈虎](#)，[JIANG Ting](#)，[ZHAO Qi-jun](#)，[CHEN Hu](#)
作者单位：[四川大学 计算机学院, 四川 成都, 610065](#)
刊名：[计算机技术与发展](#)
英文刊名：
年，卷(期)：2016(1)

引用本文格式：[江婷](#).[赵启军](#).[陈虎](#).[JIANG Ting](#).[ZHAO Qi-jun](#).[CHEN Hu](#) [基于多变量标签分布的连续型姿态估计方法](#)
[期刊论文]-[计算机技术与发展](#) 2016(1)