

# 基于级联投票机制的视频人种识别方法

刘梦源<sup>1</sup> 金永庆<sup>2</sup> 马里千<sup>1</sup> 刘宏<sup>1</sup>

(1 北京大学深圳研究生院物联网智能感知技术实验室, 广东 深圳 518055;

2 西华大学机械工程学院, 四川 成都 610039)

**摘要** 使用 LBP 算子对视频中的人脸图像提取特征, 通过线性 SVM 分类器进行人种分类, 并利用级联投票机制提高人种识别的正确率, 大大降低了视频序列中人脸误检带来的干扰. 本文提出的识别方法在 FERET 数据库中具有较好的性能; 在人种识别自行建立的 LFW 和 WEB 复杂训练数据库中通过交叉验证测试的识别率达到 91.10%; 该方法在视频数据库中的平均识别率可达 86.29%, 大量实验证明本文方法对自然场景中的光照、角度和位置变化都具有较高的鲁棒性.

**关键词** 视频; 人种识别; 识别方法; 级联投票机制; 识别率

**中图分类号** TP391 **文献标志码** A **文章编号** 1671-4512(2015)S1-0152-04

## Race classification method based on cascade voting mechanism from video

Liu Mengyuan<sup>1</sup> Jin Yongqing<sup>2</sup> Ma Liqian<sup>1</sup> Liu Hong<sup>1</sup>

(1 Lab on Intelligent Perception Engineering for WOT, Shenzhen Graduate School, Peking University, Shenzhen 518055, Guangdong China; 2 School of Mechatronic Engineering, Xihua University, Chengdu 610039, China)

**Abstract** In this paper, LBP (local binary pattern) descriptor was utilized to extract face image feature from every video by using linear SVM (support vector machine) to classify human race. Simultaneously, a cascading voting scheme was proposed to better improve the performance of race recognition task. It greatly reduces the interference caused by the proposed method, and has a prominent high recognition rate on FERET (face recognition technology) dataset. The proposed scheme can reach 91.10% on the self-established LFW (labeled faces in the wild) and WEB dataset for race recognition through cross-training method and achieve 86.29% on the video dataset. The sufficient experiments show that the algorithm are robust to illumination variation, angle and location change.

**Key words** video; race classification; classification method; cascade voting mechanism; recognition rate

人脸识别系统在众多安全领域的实际应用中扮演了重要的角色. 如公共场所的智能视频监控、人性化服务机器人、智能机器人交互技术等. 近年来国内外学者采用基于肤色、图像局部纹理特征、多特征融合等方式的传统机器学习方法来对人种进行识别分类, 虽然在特定数据库中取得了较好的检测效果, 但在真实场景下不具有鲁

棒性. 同时传统人种识别方法还过于依赖数据库的取样背景、拍摄角度等<sup>[1]</sup>, 如利用 RBF 和决策树在 FERET 数据库中的错误率达到 6%<sup>[1]</sup>; 文献[2]使用 LBP 和 WLD 算子通过 KNN 分类器在 FERET 数据库中将错误率降到 4%; 文献[3]利用 Biologically Inspired 特征区分相同性别加索人种和尼格罗人种, 获得 1.7% 的误识率<sup>[3]</sup>.

**收稿日期** 2015-06-30.

**作者简介** 刘梦源(1991-), 男, 博士, E-mail: liumengyuan@pku.edu.cn.

**基金项目** 国家自然科学基金资助项目(60875050, 60675025); 教育部博士学科点专项科研基金资助项目(20130001110011); 深圳市科技计划资助项目(JCYJ20120614152234873, 20130331144631730, 20130331144716089).

以上文献所做实验均局限于特定的数据库.文献[4]将人种识别应用于视频场景在30段视频中区分亚洲人和非亚洲人的错误率为29%;文献[5]数据没有公开,因此不具备可比性.那么如何提高针对视频中获取的自然人脸图像及复杂数据库中的识别率仍然是研究难点.针对以上存在问题本文利用级联投票机制来提高机器学习在视频以及复杂数据库中对高加索、蒙古、尼格罗多人种区分的识别率.

### 1 识别方法框架

本文根据以上问题提出人种识别系统框架如图1所示,整个识别流程分为目标人脸检测、特征提取、级联投票机制分类器三个部分.这里提出利用级联投票机制增加视频人种识别率.先对视频

中获取的每一帧图像检测并获取人脸,预处理后提取纹理特征,由于LBP具有光照不敏感性和很好的鲁棒性,因此文献[6,2,7]等使用此算法提取纹理特征.传统LBP公式定义如下

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c), \quad (1)$$

式中: $(x_c, y_c)$ 为 $3 \times 3$ 领域的中心元素,其像素值为 $i_c$ ;  $i_p$ 为区域内其他像素值; $s(x)$ 为符号函数,且

$$s(x) = \begin{cases} 1 & (x \geq 0), \\ 0 & (\text{其他}). \end{cases} \quad (2)$$

本文采用的纹理特征描述算子是由文献[8]提出的一种“等价模式”(uniform pattern)对传统LBP描述算子降维后的描述算子,并使用支持向量积线性分类器级联投票分类,通过该方法可以很大程度地降低由于复杂背景和多变人脸角度等

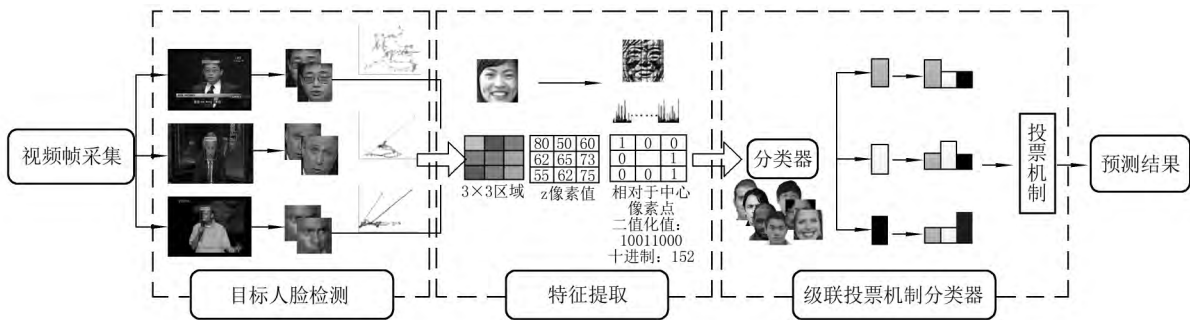


图1 基于级联投票机制的视频人种识别框图

对实验造成的影响.

### 2 级联投票机制分类方法

在获取视频每帧图片的检测目标时容易受到复杂背景的影响,将非人脸判定为人脸,目前的人脸追踪算法仍有一定误差对实验结果造成影响.本文连续获取视频每帧,同时根据物体在常态下不会跳变这一理论[4]从检测轨迹图中区分出非目标点,再利用级联投票机制降低非目标点对实验结果的影响,图2为视频序列目标检测轨迹图,在图2(a)中,非目标点 $a_i$ 与目标点群有一个明显的距离,出现 $a_i a_{i-1}$ 或 $a_{i+1} a_i$ 远远大于 $a_1 a_2, a_2 a_3, \dots$ 等目标群中的距离差,而图2(b)中则没有存在非目标点.在目标检测的过程中,由于人脸的随机移动和复杂背景等因素影响,因此在短时间内出现大量非目标点,从而影响实验对目标点的判断.

由于基于视频数据库可以对视频每帧图像处理,并可以多次对检测目标进行判断,基于这个条件利用投票机制就可以使实验结果更加可靠和准确.通过一级投票机制已经可以在很大程度上减

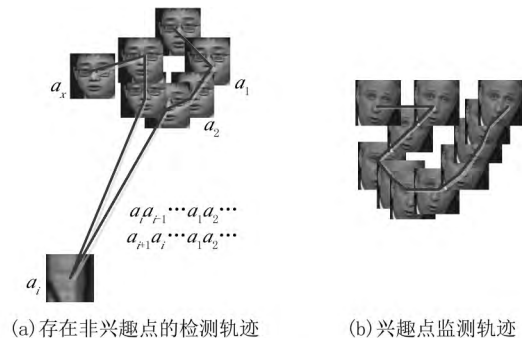


图2 视频序列目标检测轨迹图

少由于错误识别对实验判断造成的影响,但是考虑到检测目标运动状态的突变或是短时间背景造成的影响也是连续的,采用传统投票机制就可能影响最后的判断.

传统投票机制是根据视频序列中每一帧的判断结果投票得出最后结果,可以使单帧错误判断对最终判断的影响减弱.但是当连续的错误判断出现就可能使错误的比重与正确判断比重相近而导致误判,故提出在相同时间间隔内多次采用这种累加投票方式,并在检测目标不同时使用下一级投票机制对以上一级投票结果进行分类.

级联投票函数为

$$D(t) = \sum_{p=0}^P V\left(\frac{1}{T} \sum_{i=0}^T V(f(x_{t-i}))\right), \quad (3)$$

式中判定标准  $D(t)$  是根据投票函数  $V(\cdot)$  对每段的总视频帧数  $T$  每帧的分类器输出方程  $f(x_i)$  累加得出投票结果,再由投票函数  $V(\cdot)$  对  $P$  段的一级投票结果进行二级投票决定的. 这样就可以将连续误检划分到某段中而不影响对人种的分类,从而减少由于连续误检对实验结果的影响.

### 3 实验结果与讨论

#### 3.1 实验结果与讨论

本文所提的方法主要在有代表性的人脸数据库 FERET 数据库和自行建立的 LFW 及 WEB 混合数据库进行实验的证明. 其中 FERET 数据库包含高加索人种(1 210 人)、蒙古人种(410 人)、尼格罗人种(180 人);由于 FERET 数据库较为简单且数据量太少,因此本文通过网络图片(下文使用 WEB 代替)并结合 LFW 数据库,自行建立人脸姿态随意复杂的包含高加索人种(6 000 人)、蒙古人种(3 610 人)、尼格罗人种(1 290 人)的人种识别训练库;采用自行建立的 36 段多人种视频数据库.

#### 3.2 实验结果对比分析

##### 3.2.1 静态图像人种识别

本实验方法须要对自然场景中的光照、角度和位置都具有较高的鲁棒性,同时须对高维特征向量有较好的处理能力,所以本文通过比较 LBP 描述算子、Gabor 滤波特征提取算法结合 PCA 降维<sup>[9]</sup>、韦伯局部描述特征提取算法(WLD)<sup>[10]</sup>,以及各算法在 K 近邻分类器<sup>[11]</sup>、随机森林分类器、朴素贝叶斯分类器、集成分类器、判别分析器及线性 SVM 分类器<sup>[12]</sup>上的表现,测试使用的数据库为 LFW 加 WEB 网络数据库: Caucasian(4 263), Asian(1 983)和 African(734). 对图片均进行相同的归一化处理消除因实验样本不同造成的误差,然后通过改变各分类器以及描述算子的参数后得出最佳识别率,结果如表 1. 从表 1 中可以得出结论:在复杂数据库中 LBP 算子和线性 SVM 分类器对高维特征向量处理具有较好的辨识度和速度,同时实验数据也验证了 LBP 具有对光照不敏感和对纹理特征处理的鲁棒性,SVM 分类器处理高维特征时具有较好的速度和鲁棒性.

为了验证 LBP 算子以及线性 SVM 分类器的鲁棒性,这里分别在 LFW,WEB 和 FERET 数

表 1 描述算子在分类器中的识别率 %

分类器	LBP	Gabor	WLD
KNN	56.82	49.24	33.33
Random Forest	61.36	56.82	59.09
Native Bayes	45.45	52.27	48.48
Ensembles	44.70	43.93	56.06
Discriminant	68.94	67.42	59.85
SVM	80.56	67.42	62.88

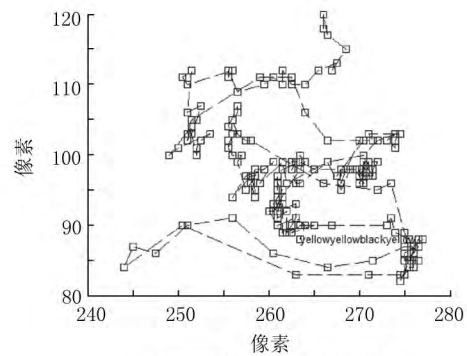
据库中进行识别方法的交叉验证. 如表 2 所示,均为 10 折交叉验证,通过比较数据,可以看出该方法在 FERET 数据库中的识别率相对较低,造成原因可能是数据量太少或数据分布不均匀;并不能满足实验数据库对复杂度的要求,所以最终实验采用 LFW 和 WEB 数据库进行视频人种分类器训练.

表 2 数据库比较

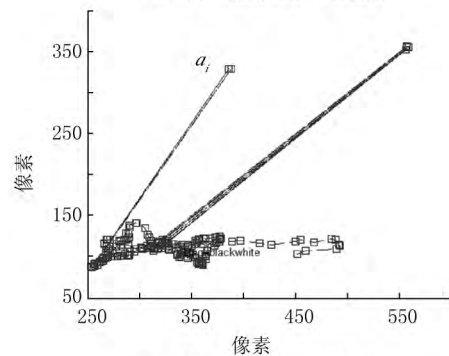
数据库	人种			识别率/%
	高加索	蒙古	尼罗格	
LFW,WEB	6 000	2 610	1 290	91.10
FERET	1 210	410	180	89.75

##### 3.2.2 相同检测目标判断

实际检测过程中使用检测脸部的中心位置代替检测目标位置,本实验中每段视频长度为 10 s 左右,38 帧/s,并每 120 帧进行一次一级投票,多次检测为非人脸或视频结束时综合分类检测目标. 如图 3(a)为同一检测目标的运动轨迹,轨迹



(a) 同一检测目标运动轨迹



(b) 非同同一检测目标运动轨迹

图 3 视频序列目标检测轨迹图

中没有孤立点.图 3(b)则存在明显的非目标点  $a_i$  和  $a_j$ ,该孤立点则会对实验结果造成影响;实际实验中出现运动轨迹和原来轨迹、运动范围有明显区别也判定为非相同目标.

### 3.2.3 识别方法的识别率

使用训练好的分类器对 36 段视频分别测试.表 3 给出了对人种识别的统计结果,其中 1(y)/2(b)表示 1 次判断为 Asian(yellow),2 次判断为 African(black);3(w)表示 3 次判断为 Caucasian (white); $W_1$  和  $W_2$  表示高加索人种, $Y_1$  和  $Y_2$  表示蒙古人种, $B_1$  和  $B_2$  表示尼格罗人种, $S_1 \sim S_6$  为测试组.表 4 是采用无级联与级联两种识别方法投票得出的分类识别率对比.由表 4 可知:级联投票的平均识别率(86.29%)明显高于无级联投票的平均识别率(82.96%).

表 3 分类统计结果

人种	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$
$W_1$	3(w)	2(w)/ 1(y)	3(w)	3(w)	2(w)/ 1(b)	3(w)
$W_2$	3(w)	3(w)	3(w)	3(w)	3(w)	3(w)
$Y_1$	1(y)/ 2(b)	0(y)/ 3(w)	3(y)	3(y)	3(y)	3(y)
$Y_2$	3(y)	2(y)/ 1(b)	3(y)	0(y)/ 3(w)	3(y)	3(y)
$B_1$	3(b)	3(b)	3(b)	3(b)	3(b)	3(b)
$B_2$	3(b)	0(b)/ 3(w)	3(b)	0(b)/ 3(w)	2(b)/ 1(y)	2(b)/ 1(w)

表 4 两种识别方法识别率对比 %

人种	无级联	级联投票
高加索	94.44	100
蒙古	75	75
尼格罗	78.88	83.33

本文针对在复杂数据库或视频中人种识别会受自然场景、光照等影响这一问题,提出了级联投票机制的视频人种识别方法,并建立具有一定复杂性的可靠数据库.实验数据表明:该识别方法在 FERET 代表性数据库中具有很不错的效果;在复杂数据库中,对多分类的人种识别具有很高的识别率;同时在视频数据库中也有较好的分类效果.视频特征和图片特征有很大差别,本文未来将针对视频特征对实验进行进一步的对比和改进;同时建立该领域具有客观性的数据库.

### 参 考 文 献

[1] Gutta S, Wechsler H, Phillips P J. Gender and ethnic classification of face images[C]// Proceedings of

the 3rd International Conference on Face & Gesture Recognition. Nara: IEEE, 1998: 194-199.

[2] Muhammad G, Hussain M, Alenezy F, et al. Race classification from face images using local descriptors [J]. International Journal on Artificial Intelligence Tools, 2012, 21(5): 44-51.

[3] Guo Gudong, Mu Guowang. A study of large-scale ethnicity estimation with gender and age variations [C]// Proc of Computer Vision and Pattern Recognition Workshops. San Francisco: IEEE, 2010: 79-86.

[4] Shakhnarovich G, Viola P A, Moghaddam B. A unified learning framework for real time face detection and classification [C]// Proceedings of the 3rd International Conference on Face & Gesture Recognition. Washington: IEEE, 2002: 14-21.

[5] Demirkus M, Garg K, Guler S. Automated person categorization for video surveillance using soft biometrics [C]// Proc of SPIE Defense, Security, and Sensing. Florida: SPIE, 2010: 1-12.

[6] Klare B F, Burge M J, Klontz J C, et al. Face recognition performance: role of demographic information [J]. Information Forensics & Security IEEE Transactions on, 2012, 7(6): 1789-1801.

[7] Salah S H, Du H, Aljawad N. Fusing local binary patterns with wavelet features for ethnicity identification [C]// Proc of 20th International Conference on Systems, Signal Image Process. Bucharest: IEEE, 2013: 416-422.

[8] Ojala T, Pietkainen M, Harwood D. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition, 1996, 29(1): 51-59.

[9] Lin Hui, Lu Huchuan, Zhang Lihe. A new automatic recognition system of gender, age and ethnicity [C]// Proc of 6th World Congress on Intelligent Control and Automation. Dalian: IEEE, 2006: 9988-9991.

[10] Muhammad G, Hussain M, Alenezy F, et al. Race recognition from face images using Weber local descriptor [C]// Proc of 19th International Conference on Systems, Signals and Image Processing. Vienna: IEEE, 2012: 421-424.

[11] Khaung Tin H H, Myint Sein M. Race identification for face images [J]. International Journal on Information Technology, 2011, 01(2): 35-35.

[12] Manesh F S, Ghahramani M, Tan Y P. Facial part displacement effect on template-based gender and ethnicity classification [C]// Proc of 11th International Conference on Control Automation Robotics & Vision. Singapore: IEEE, 2010: 1644-1649.