

文章编号: 1000-5013(2010)01-0032-05

# 角度与动态时间归一化的步态识别

陈璇, 吴清江, 路远

(华侨大学 计算机科学与技术学院, 福建 泉州 362021)

**摘要:** 首先选择具有鲁棒性的累积角度特征作为步态特征, 并对每个序列提取角度特征, 构成特征矩阵. 根据步态序列具有线性特征的特点, 在提取步态周期的基础上, 采用动态时间归一化做序列匹配, 计算最终的特征距离. 实验结果表明, 算法具有快速、稳健的特征, 并且在 120 人的步态数据库 0(9) 与 90(9) 的视角上取得较好的识别率.

**关键词:** 步态识别; 角度特征; 动态时间归一化; 距离计算

**中图分类号:** TP 391.4

**文献标识码:** A

人体步态由许多因素决定, 包括体格、体重、鞋子高度、地面、衣着等. 由于人们经常改变衣着、行走速度, 以及摄像机拍摄角度差异, 使步态识别成为一个具有挑战性的课题. 王亮等<sup>[1]</sup>将轮廓边缘展开, 在特征空间投影并用 NN 做分类. 黄建新等<sup>[2]</sup>利用 Kalman 滤波器和自适应背景模型对背景建模, 在人体跟踪上取得较好的效果. 陈彦<sup>[3]</sup>对图像序列进行 PCA 运算, 得到 PC 子空间, 对 PC 系数进行 ICA 运算, 并计算周期序列的平均 IC 系数, 对平均 IC 系数以马氏距离为标准用最近邻法做识别. 柴艳妹等<sup>[4]</sup>首先将检测出的二维人体侧影分为头部、躯干和腿部 3 个区域, 分别提取每个区域的目标面积, 将这些面积特征和人体的宽高比特征一起构成步态特征矢量用于训练和识别. 田见光等<sup>[5]</sup>利用步态轮廓图像边界到重心的距离矢量对步态轮廓图像进行描述, 采用步态图像的高宽比进行步态的准周期性分析, 利用隐马尔可夫模型进行步态时变数据匹配识别. 赵国英等<sup>[6]</sup>提出了分形维数小波分析描述子步态识别算法, 利用训练信号, 得到广义 M 进制多尺度滤波器, 使得重构信号对原始信号达到最佳逼近. 并通过最优滤波器提取特征, 进行模式分析. 本文提出一种具有较强鲁棒性的基于累积角度的步态特征.

## 1 步态特征提取

### 1.1 步态检测与跟踪

采用基于色度坐标的高斯混合模型<sup>[7-8]</sup>提取背景图像. 首先, 将 RGB 通道颜色转换为色度坐标, 即

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}, \quad b = \frac{B}{R + G + B}. \quad (1)$$

另外, 还需要加入亮度信息, 有

$$I = R + G + B. \quad (2)$$

$I$  的取值范围为  $[0, 255]$ , 并且使  $r, g$  取值范围为  $[0, 255]$ , 即  $r = R / (R + G + B) \times 255$ ,  $g = G / (R + G + B) \times 255$ , 从而构成特征空间  $(r, g, I)$ .

混合高斯模型的基本思想: 对每一个像素点, 定义  $K$  个状态来表示其所呈现的颜色,  $K$  个状态中每个状态用一个高斯函数表示. 这些状态一部分表示背景的像素值, 其余部分则表示运动前景的像素值. 若每个像素点颜色取值用变量  $x_t$  表示, 其概率密度函数可用  $K$  个三维高斯函数表示为

收稿日期: 2008-10-19

通信作者: 吴清江(1949-) 男, 教授, 主要从事语音识别、人脸识别和步态识别的研究. E-mail: wuqingjiang@msn.com.

基金项目: 福建省自然科学基金资助项目(2006J0036)

$$P(x_t) = \sum_{i=1}^K \frac{w_{i,t}}{(2\pi)^{n/2} |\Sigma_{i,t}|} \exp[-\frac{1}{2}(x_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (x_t - \mu_{i,t})], \quad i = 1, 2, \dots, K. \tag{3}$$

式(3)中:  $(x, \mu_{i,t}, \Sigma_{i,t})$  是  $t$  时刻的第  $i$  个高斯分布,  $\mu_{i,t}$  为第  $i$  个高斯分布在  $t$  时刻的均值,  $\Sigma_{i,t}$  为协方差矩阵,  $w_{i,t}$  为权重, 且有  $\sum_{i=1}^K w_{i,t} = 1$ ,  $n$  为  $x_t$  的维数.

出于计算量的考虑, 一般认为各个像素点的色彩通道都是相互独立的, 并且具有相同的方差, 其协方差矩阵为

$$\Sigma_{i,t} = \frac{\sigma^2}{k} \mathbf{I}. \tag{4}$$

把各个高斯分布按照  $w_{i,t} / \Sigma_{i,t}$  的值从大到小排列, 其值越大, 排的顺序越靠前. 若  $K=3$ , 则  $w_{i,t} / \Sigma_{i,t}$  值最大的高斯分布则认为是背景; 而在  $K>3$  的情况下, 若前  $M_1$  个分布.

1.2 特征描述<sup>[9]</sup>

以图像中心为原点, 重新建立坐标系, 对应关系为  $(x, y) = (s_y - c_y, c_x - s_x)$ . 选择某个角度范围内, 像素点到原点的距离为特征, 其角度转换为

$$A(i) = \frac{1}{\sum_{k=1}^k N_{F,i}} \sum_{(x,y) \in F,i} I(x,y) \sqrt{x^2 + y^2}. \tag{5}$$

式(5)中:  $F$  为以原点为中心, 角度为  $(-\pi/2, +\pi/2)$  范围的区域的前景像素点,  $N_{F,i}$  为第  $i$  帧在  $F$  区域内像素点个数. 式(5)表示从第 1 帧到第  $k$  帧前景像素点的总长度除以前景像素点的总个数, 得到前景像素点的平均长度, 即可以得到前景像素点与剪影中心的平均距离, 如图 1 所示. 步态剪影的转换都与步态向量的维数  $K=360/\Delta\theta$  有关. 其中:  $\Delta\theta$  为角度的间距大小. 此处, 取  $5^\circ$ . 一帧图像的步态的特征向量可以表示为

$$A = [A(0), A(0), \dots, A(K-1)]. \tag{6}$$

所有的特征向量都执行了大小归一化操作, 最大的特征值为  $M$ , 文中  $M$  取 255. 缩放操作使式(6)具有缩放不变性. 这个特性较常用, 因为人们很可能会以不同的角度或不同的距离在摄像头前行走. 整个步态序列用一系列的上述形式的向量表示. 其特征矩阵为

$$V = \begin{bmatrix} A_0(0) & A_1(0) & \dots & A_{L-1}(0) \\ A_0(1) & A_1(1) & & A_{L-1}(1) \\ \dots & & & \dots \\ A_0(K-1) & A_1(K-1) & \dots & A_{L-1}(K-1) \end{bmatrix}. \tag{7}$$

式(7)中:  $A_i(k)$  为第帧的第  $k$  个角度转换范围,  $L$  为步态序列的帧数量. 步态序列的转换表达可根据  $FV_i = \text{sum}(A_i)$ ,  $FV = [FV_0, \dots, FV_i, \dots, FV_{L-1}]$  计算出.

2 步态识别

对于同一种运动, 不同人或相同人在不同时间下, 其运动速度和长度也不同. 如果对于其中的每个动作, 都在参考序列中根据最近距离匹配, 往往会出现本应在后面的动作匹配到前面的情况, 造成运动的抖动. 处理这种序列匹配的问题, 通常采用基于线性匹配的线性时间归一化(LTN)和基于非线性匹配的动态时间变形(DTW)两种方法. 实验证明, DTW 方法在语音识别领域是效果最好的序列匹配方法, 然而大量的文献表明, 步态在周期内是线性的运动. 因此, 采用 LTN 处理步态的序列匹配.

2.1 线性时间归一化

一种有效的计算训练周期和测试周期距离长度的方法是执行时间归一化操作. 即需要通过训练集(测试集)中某周期的每个特征向量与它相应的测试集(训练集)某周期的特征向量距离相加累计, 以得到训练集(测试集)中该周期的特征向量距离. 采用线性时间归一化, 即不同长度周期的帧之间的联系通过一个线性规则计算得到, 如图 2 所示.

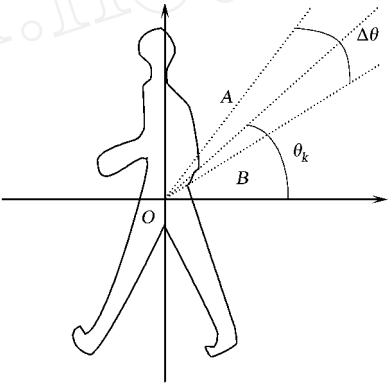


图 1 角度转换的图形表达  
Fig.1 Graphic of angle transfer

特征距离是通过测试集的所有周期 $\{I_1^p, I_2^p, \dots, I_I^p\}$ 与训练集的所有周期 $\{I_1^g, I_2^g, \dots, I_J^g\}$ . 其中: $I, J$  分别为测试集和训练集的周期数; $I_i^p, I_j^g$  分别为第  $i$  个测试周期和第  $j$  个训练周期; $T_i$  与  $R_j$  分别为第  $i$  个测试集和第  $j$  个训练集周期的帧数量; $t, r$  分别为测试集和训练集的特征标号; $T_r, R_t$  为用线性规则计算得出的标号集合. 两个步态周期之间的距离为

$$d_{i,j} = \frac{1}{2} \left[ \sum_{r=1}^{T_i} \min_{t \in R_t} d_e(FV_t, FV_r) + \sum_{t=1}^{T_i} \min_{r \in R_t} d_e(FV_t, FV_r) \right] \quad (8)$$

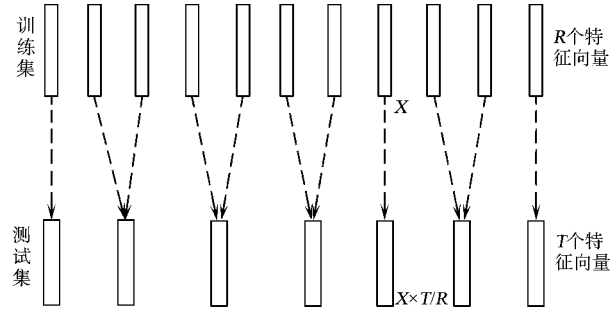


图2 线性时间归一化过程

Fig. 2 Graphic of linear time normalization

式(8)中: $i = 1, \dots, I; j = 1, \dots, J$ . 由此,可以得到一个  $I \times J$  的矩阵为

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,J} \\ d_{2,1} & d_{2,2} & \dots & d_{2,J} \\ \dots & \dots & \dots & \dots \\ d_{I,1} & d_{I,2} & \dots & d_{I,J} \end{bmatrix} \quad (9)$$

式(9)可用于计算训练集与测试集距离的最终结果.

## 2.2 最终距离计算

计算出测试周期和训练周期之间的距离矩阵之后,需要找到一种方法计算测试序列和训练序列之间的距离  $D_r$ ,以描述两个步态序列之间的相似度.  $D_r$  不论是对于训练集还是测试集来说,都应该是对称的. 整个最终距离的计算过程,如图3所示.

首先,需要找到矩阵  $D$  每行和每列的最小值. 即

$$\left. \begin{aligned} d_{\min,i}^p &= \min_j \{d_{i,j}\}, & i &= 1, \dots, I; \\ d_{\min,i}^g &= \min_i \{d_{i,j}\}, & j &= 1, \dots, J. \end{aligned} \right\} \quad (10)$$

其次,计算出的每行的最小值和每列的最小值,可以形成两个集合为

$$\left. \begin{aligned} J_p &= \{d_{\min,1}^p, \dots, d_{\min,I}^p\}, \\ J_g &= \{d_{\min,1}^g, \dots, d_{\min,J}^g\}. \end{aligned} \right\} \quad (11)$$

第三,测试序列或训练序列的一些周期可能在识别个人身份的时候不是很具有代表性,因此,加入这些周期计算的特征距离可能会偏大或偏小. 为了保证特征距离  $D_r$  的准确性,要把这些不具代表性的周期去除. 鉴于以上的结论,分别采用集合  $J_p$  与  $J_g$  最小值,有

$$D_r^p = \min \{d_{\min,1}^p, \dots, d_{\min,I}^p\}, \quad D_r^g = \min \{d_{\min,1}^g, \dots, d_{\min,J}^g\}. \quad (12)$$

最后,对它们取平均值,以计算测试序列与训练序列的最终距离. 即

$$D_r = \frac{1}{2} (D_r^p + D_r^g). \quad (13)$$

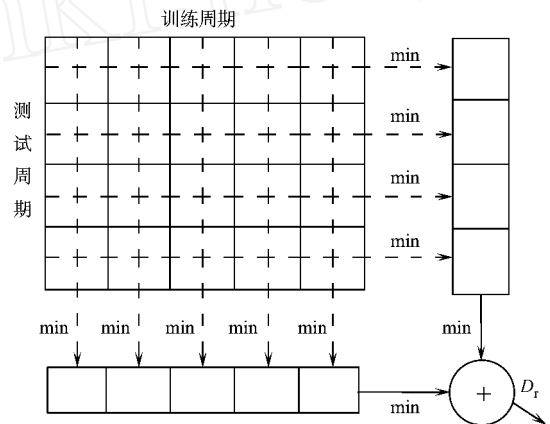


图3 最终距离计算过程

Fig. 3 Graphic of final distance counting

## 3 实验验证

### 3.1 数据来源

采用 CASIA :Dataset B 步态数据库<sup>[4]</sup>. 该数据库采集于 2005 年 1 月,是一个大规模的、多视角的步态库. 视频以 mjpeg 格式编码,帧图像原始大小为 320 px × 240 px,以每秒 25 帧的速度同时从 11 个角度(0°, 18°, 36°, ..., 180°)拍摄. 每个视频序列包括 2~3 个步态周期. 该数据库有 124 个人的步态,包括 93 个男性和 31 个女性,其中亚洲人 123 个和欧洲人 1 个. 拍摄对象都是 20~30 岁的年轻

人. 每个人在正常状态下走 6 次,穿上大衣后走 2 次,带包状态下走 2 次. 因此,该数据库总共包括  $10 \times 11 \times 124 = 13\,640$  个视频序列.

3.2 基于角度特征的 LTN 识别率

提取步态剪影,并采用角度特征作为步态特征,用 LTN 做步态分析,结果如表 1 所示. 由表 1 可知,在视角( $\theta$ )为  $90^\circ$  的情况下,识别率( $r$ )可达到  $91.9\%$ ;在视角为  $0^\circ, 180^\circ$  时,识别率可达  $83\%$  以上;而当斜侧面行走时,识别率偏低,为  $66\%$  左右.

表 1 基于角度特征的 LTN 识别率

Tab. 1 LTN recognition using angle feature											
$\theta / (^\circ)$	0	18	36	54	72	90	108	126	144	162	180
$r / \%$	83.9	83.1	67.7	73.4	87.1	91.9	88.7	72.6	66.1	81.5	84.7

3.3 分类性能度量

分类性能度量(Rank Order Statistic, ROS)首先在 FERET 脸像识别算法的评估协议中被提出. 它定义为一个测试度量的实际类别,在它的最前  $k$  个匹配值之间的累积概率  $p(k)$ . 性能统计特性以累积匹配分值( $C$ )来报告,阶次  $k$  画在横轴上,而垂直轴是正确匹配的累计百分比( $r$ ). 在 120 个人,普通状态,6 个角度的条件下,使用 ROS 评估算法的性能,评测结果如图 4 所示.

从图 4 可以知道,角度为  $108^\circ \sim 180^\circ$  的 ROS 匹配结果与  $0^\circ \sim 72^\circ$  的相似;关于  $90^\circ$  的匹配结果对称,即  $108^\circ$  与  $72^\circ$  的匹配度类似; $126^\circ$  与  $54^\circ$  的匹配度类似,依次类推. 实验结果表明, $0^\circ, 18^\circ, 72^\circ, 90^\circ$  的匹配度比较好,而  $36^\circ, 54^\circ$  的匹配度则不大理想. 即侧面行走时匹配度不是很好,分类效果不是很好.

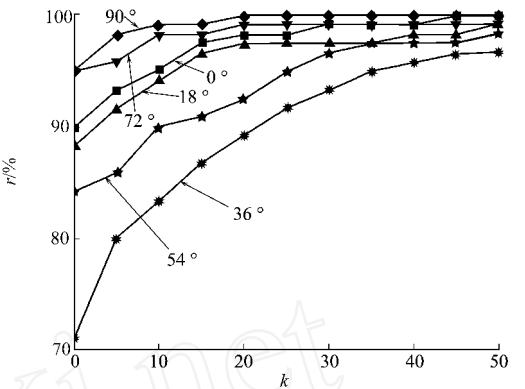


图 4 分类性能度量评估结果

Fig. 4 Result of rank order statistic

3.4 识别方法

把使用 LTN 进行序列匹配之后得出的识别率,与直接用最近邻规则(NN)计算出的识别率进行比较,结果如表 2 所示. 从表 2 中可以看出,使用 NN 分类的结果均为  $70^\circ$  左右,变化不大;而使用 LTN 做序列匹配之后,识别率波动比较大,在  $0^\circ, 90^\circ, 180^\circ$  处识别率最高, $36^\circ$  与  $144^\circ$  识别率最低.

使用 LTN 进行序列匹配后,其识别率总体来说比没有进行序列匹配直接用 NN 得到的识别率要好,仅仅在  $36^\circ$  与  $144^\circ$  时,比 NN 方法识别率略低. 特别是在  $90^\circ$  视角时,使用 LTN 序列匹配之后的结果比 NN 的识别率要高  $20\%$ .

表 2 LTN 与 NN 识别率的比较

Tab. 2 Recognition rate between LTN and NN											
$\theta / (^\circ)$	0	18	36	54	72	90	108	126	144	162	180
NN / %	79.9	79.0	72.6	74.2	76.6	71.0	70.2	73.4	70.2	80.6	82.3
LTN / %	83.9	83.1	67.7	73.4	87.1	91.9	88.7	72.6	66.1	81.5	84.7

3.5 不同算法

目前,比较多的识别率是 25 人的小数据库上实验得出的. 因此,在 25 人的数据库上,将提出的基于角度特征与 LTN 的步态识别方法,以及与文[3-6]中方法的识别率进行了对比,结果如表 3 所示. 其中,数据集选定为 25 人,视角为  $90^\circ$ ,并加入速度影响因素下的识别率.

从表 3 的数据可知,提出的方法在行走速度影响上,识别率高于另外几种算法.

表 3 几种不同算法识别率的比较

Tab. 3 Recognition rate among several method							
方法	数据集/人	影响因素	$r / \%$	方法	数据集/人	影响因素	$r / \%$
文[3]	25	行走速度	64.0	文[6]	25	行走速度	83.33
文[4]	25	行走速度	68.0	本文	25	行走速度	91.0
文[5]	25	行走速度	81.0				

3.6 不同状态

目前,国内外研究比较多的是基于普通状态下的步态识别,即没有加入太多对外形轮廓形象很大的因素.通过对普通、大衣、背包状态下的识别率标胶,结果如表 4 所示.

从表 4 可知,由于大衣与背包对步态的外形影响很大,会造成重心下移,轮廓变形,前景像素增多,因此识别率较普通状态有较大下降.

表 4 各个不同状态下的识别率

Tab.4 Recognition rate among different states											%
/ (°)	0	18	36	54	72	90	108	126	144	162	180
普通	83.9	83.1	67.7	73.4	87.1	91.9	88.7	72.6	66.1	81.5	84.7
大衣	27.4	16.1	15.3	26.6	32.3	35.5	33.9	29.0	18.5	19.3	29.0
背包	25.0	17.7	16.9	24.2	29.0	36.7	30.6	25.8	17.7	18.5	26.6

4 结束语

提出了一种基于角度与动态时间归一化的步态识别方法.实验结果表明,该方法在较大步态数据库的各个角度上均取得较好的识别率.对外形轮廓影响比较大的两种状态,需要做进一步的研究,这对特征的鲁棒性提出了非常高的要求.此外,遮挡、场景的一些干扰也将是今后工作的一个研究方向.

参考文献:

[1] 王 亮,胡卫明,谭铁牛.基于步态的身份识别[J].计算机学报,2003,26(3):353-360.

[2] 黄建新. Kalman 滤波的人体运动位置跟踪算法[J].华侨大学学报:自然科学版,2003,24(3):254-256.

[3] 陈 彦.基于独立分量分析的步态识别方法研究[D].西安:西安电子科技大学,2007

[4] 柴艳妹,赵荣椿.一种新的基于区域特征的快速步态识别方法[J].中国图象图形学报,2006,11(9):1260-1265.

[5] 田光见,赵荣椿.基于连续隐马尔可夫模型的步态识别[J].中国图象图形学报,2006,11(6):867-871.

[6] 赵国英.基于视频的步态识别[D].北京:中国科学院,2005.

[7] MITTAL A,PARAGIOS N. Motion-based background subtraction using adaptive kernel density estimation[C] Proc of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Princeton:[s. n. ],2004:302-309.

[8] POWER P,SCHOONEES J. Understanding background mixture models for foreground segmentation[C] Proc of Image and Vision Computing. Auckland:[s. n. ],2002:267-271.

[9] BOU GOURIS N,PLATANIOTIS K,HA TZINA KOS D. An angular transform of gait sequences for gait assisted recognition[C] Proc of 2004 International Conference on Image Processing,2004,2(24/27):857-860.

Gait Recognition of Angle and Dynamical Time Normalization

CHEN Xuan , WU Qing-jiang , LU Yuan

(College of Information Science and Engineering , Huaqiao University , Quanzhou 362021 , China)

**Abstract :** Firstly , robust accumulative angle feature is chosen to be the basic gait feature , and the angle feature of sequences are saved as matrixes. Secondly , on the basis of gait sequence linearity , extract the gait cycle , and then introduce dynamical time normalization into sequence matching , to get the final feature distance. Experimental result shows that the proposed algorithm performs an encouraging recognition rate with relatively lower computational cost at 0 (°) and 90 (°) viewing angle in large gait dataset for 120 persons.

**Keywords :** gait recognition ; angle feature ; dynamical time normalization ; distance calculation

(责任编辑: 鲁 斌 英文审校: 吴逢铁)

