

基于概率模型的行人四肢 自遮挡的检测

于雪松¹ 刘家锋¹ 唐降龙¹ 黄剑华¹

摘要 针对人体运动跟踪领域中的自遮挡现象,提出了一种基于概率模型的行人四肢自遮挡的检测算法。首先,该方法定义了基于马尔可夫模型的自遮挡状态概率模型;其次,利用椭圆肤色模型,定义了层次肤色模型;最后,通过上述模型,算法将行人四肢自遮挡状态的识别转换为计算自遮挡状态转换概率的过程。实验表明,该方法具有较高的准确性。

关键词 自遮挡, 马尔可夫模型, 椭圆肤色模型, 层次肤色模型

DOI 10.3724/SP.J.1004.2010.00610

Estimating the Pedestrian 3D Motion Indoor via Hybrid Tracking Model

YU Xue-Song¹ LIU Jia-Feng¹ TANG Xiang-Long¹
HUANG Jian-Hua¹

Abstract Focusing on the problem of self-occlusion in the field of human motion tracking, the paper deals with an algorithm for detecting the pedestrian limbs self-occlusion probability model. First, the algorithm defines the self-occlusion state probability model based on a Markov model. Second, the paper proposes a hierarchical skin model using ellipse skin model. Last, the algorithm changes the detection of pedestrian limbs self-occlusion to the calculation of the self-occlusion state transition probability. The result of experiment shows that the algorithm has a higher accuracy.

Key words Self-occlusion, Markov model, ellipse skin model, hierarchical skin model

近年来,人体运动跟踪逐渐成为计算机视觉领域中最活跃的研究课题之一。在文献[1-2]中, Moeslund 等对人体运动跟踪作出了全面的总结。

从文献[1-2]中可知,轮廓^[3]、边缘^[4]、颜色直方图^[5]等图像特征被用于人体运动跟踪。然而,在人体运动过程中,由于肢体之间自遮挡现象而导致的轮廓特征的二义性问题以及其他特征丢失严重影响了跟踪结果和精度。因此,在人体跟踪领域中,人体自遮挡问题已经成为亟需解决的问题。

在现有的文献中,多摄像机系统^[6-9]以及基于学习的方法^[10-12]被广泛地应用以解决自遮挡问题。但是,高昂的造价、庞大的设备、安装调试困难等缺点,使得基于多摄像机的方法难以得到普遍应用。在基于学习的方法中,流形(Manifold)^[10-11]被用来对人体运动状态参数空间进行降维处理,高斯过程本征变量模型(Gaussian process latent variable model, GPLVM)^[12]等方法被用来进行样本训练,然而,这些方法存在着大量训练样本的需求、训练时间开销过大等问题。

本文对室内人体行走视频图像进行了统计分析。相对于室外自然场景,室内人体运动具有以下特点:1)良好光照条

件;2)几乎没有任何外部遮挡,因此,肤色特征比较容易提取;3)在本文所选取的实验视频中,目标人体的四肢均有部分裸露的皮肤。经过分析,可以得出:随着人体肢体之间自遮挡状态的不同,肤色特征具有明显的规律性。

针对上述问题,本文提出了一种行人四肢自遮挡的检测方法。该方法将步行人体运动按照人体四肢自遮挡的情况分为四种自遮挡状态,利用马尔可夫模型,将人体自遮挡的检测转换为计算人体运动自遮挡状态转换概率的问题,从而降低了时间开销和设备成本。本文的第1节定义了步行人体运动自遮挡情况的四种状态,并给出了基于马尔可夫模型的自遮挡状态转换模型;第2节定义了层次肤色模型,并介绍了相应的匹配算法;第3节给出了算法应用于 HumanEva 人体运动实验数据集的实验结果,并进行相关的分析;最后为结论。

1 自遮挡状态概率模型

1.1 自遮挡分析

由于室内步行人体的运动并不包含甩臂等剧烈运动情况,并且在通常情况下,步行人体运动跟踪也很少考虑甩臂等剧烈运动情况,因此,根据正常步行人体运动的特点,本文将步行人体运动姿态按照四肢自遮挡情况分为四种状态:前行状态(Front walking, FW),左侧遮挡状态(Left self-occlusion, LSO),背行状态(Back walking, BW),右侧遮挡状态(Right self-occlusion, RSO),见图1。按照从左至右的顺序,图1显示了这四种状态下步行人体运动姿态。

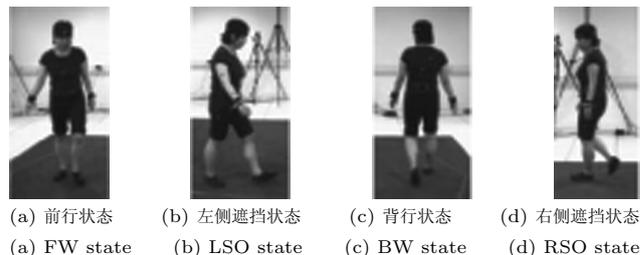


图1 步行人体运动自遮挡状态示意图

Fig. 1 The diagram for occlusion state of pedestrian walking

根据图1,结合步行人体运动的先验知识,在这四种自遮挡状态中,面部、手臂和腿部呈现出不同的特征细节。如在图1(a)和图1(c)中,两只手臂之间重心距离较大,而在图1(b)和图1(d)中有一只手臂已经被行人本身自遮挡。在图1(a)中的面部肤色面积明显要比其他图像中的面部肤色面积大,而图1(c)中的面部已经完全被遮挡。图2为对应图1的图像中利用椭圆肤色模型检测出来的肤色特征,并按面部、手臂和腿部进行分组排列,其中,图2(a)为相应的面部肤色特征,图2(b)为相应的手臂肤色特征,图2(c)为相应的腿部肤色特征。

为了便于分析,我们对样本图像进行手工标注步行人体运动状态。通过对标注后的样本进行特征提取,本文对所提取出来的特征按照不同的自遮挡状态进行分析,分析结果见图3。其中,图3(a)为面部皮肤面积在不同自遮挡状态下的变化曲线,图3(b)为两臂重心距离在不同自遮挡状态下的变化曲线。根据图3,我们可以得出不同自遮挡状态下的皮肤特征有着较为明显差别的结论。

收稿日期 2009-03-03 录用日期 2009-06-24
Manuscript received March 3, 2009; accepted June 24, 2009
国家自然科学基金(60672090)资助
Supported by National Natural Science Foundation of China (60672090)
1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001
1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

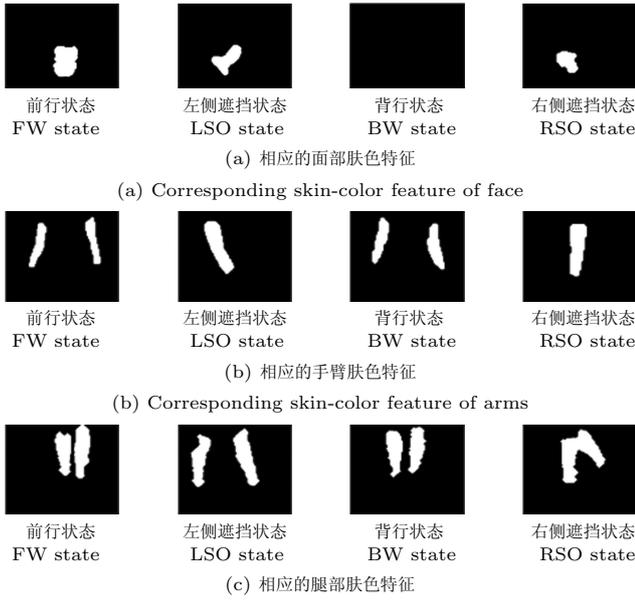


图2 不同状态下相应的肢体部位肤色特征

Fig. 2 Skin-color features of limbs on different states

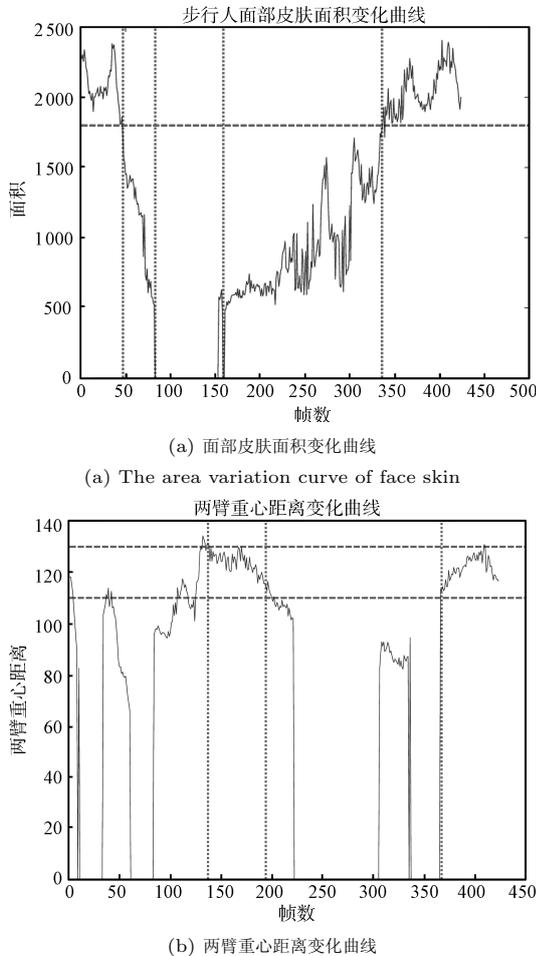


图3 行人皮肤部分特征变化曲线

Fig. 3 The curve for some skin features of pedestrian walking

1.2 自遮挡概率模型

在 HumanEva^[13] 数据集所拍摄步行人体运动视频中, 考虑到人体运动的连续性, 任意时刻的人体姿态都可由前文所定义的四种自遮挡状态来表示, 相应的人体运动状态转移过程见图 4. 在图 4 中, 排除了 4 种不可能的人体运动状态转移: 1) FW → BW: 由前行状态直接转换到背行状态; 2) BW → FW: 由背行状态直接转换到前行状态; 3) LSO → RSO: 由左侧遮挡状态直接转换到右侧遮挡状态; 4) RSO → LSO: 由右侧遮挡状态直接转换到左侧遮挡状态. Moeslund 等^[1-2] 认为, 人体运动可以用一阶马尔可夫过程来解释, 每一时刻的状态由上一时刻的状态以及当前时刻的观察值所约束. 结合自遮挡状态的定义, 则有如下结论: 任意时刻的人体自遮挡状态由上一时刻的状态和当前时刻的观测值所约束. 于是, 自遮挡状态的检测就可以转化为利用马尔可夫模型对当前时刻对自遮挡状态的寻优过程.

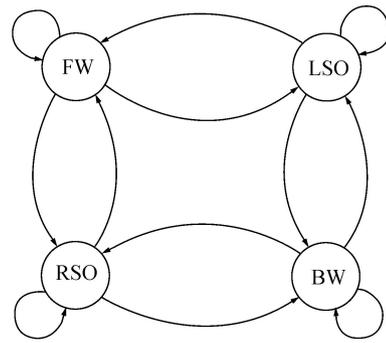


图4 基于自遮挡状态的人体运动状态转移图

Fig. 4 Pedestrian motion states transition diagram based on self-occlusion states

令 $O = \{O_1, O_2, O_3, O_4\}$, 按照下标升序分别对应前行状态 (FW)、左侧遮挡状态 (LSO)、背行状态 (BW) 和右侧遮挡状态 (RSO). 令 $B^A = \{C_C^A, C_A^A, C_D^A\}$ 表示手臂皮肤轮廓特征集, 其中 C_C^A 表示所检测到的手臂皮肤轮廓个数, 即手臂个数; C_A^A 表示所检测到的手臂皮肤轮廓重心之间的连线与水平轴的夹角; C_D^A 表示手臂皮肤轮廓重心之间的欧氏距离. 参照手臂皮肤轮廓特征集的定义, 令 $B^L = \{C_C^L, C_A^L, C_D^L\}$ 表示腿部皮肤轮廓特征集. B^A, B^L 之间相互独立, C_C^A, C_A^A 之间相互独立, 同样, C_D^A, C_D^L 之间相互独立. 根据 HumanEva 数据集^[13] 中视频图像的特点, 上述特征可按照第 2.1 节中所描述的方法获得.

定义 O 为自遮挡状态集合, 定义 $B = \{B^A, B^L\}$ 为观测状态集合, 则自遮挡状态随时间变化的序列 S 可定义为

$$S = \{s_1, s_2, \dots, s_t\} \tag{1}$$

其中, $s_t \in O(i), 1 \leq i \leq 4$. 同样可以定义观测状态 B 随时间变化的序列 Z 为

$$Z = \{z_1, z_2, \dots, z_t\} \tag{2}$$

其中, $z_t \in B, 1 \leq t \leq T$.

根据以上定义以及马尔可夫模型, 在任意时刻 t , 状态 s_t 的最大后验概率 (Maximum a posteriori, MAP) 可通过下式计算获得:

$$P(s_t|z_t) = \max_{s \in O} P(s|s_{t-1})P(z_t|s) \tag{3}$$

其中, $s \in O$ 表示人体运动的四种自遮挡状态; $P(s|s_{t-1})$ 表示由时刻 $t-1$ 的状态到当前时刻状态的转移概率; $P(z_t|s)$ 表示在当前状态下得到观测值 z_t 的概率.

在任意时刻 $t \geq 1$ 时, 式 (3) 需要计算全部四种自遮挡状态的概率. 然而, 如前文所述, 在任意一种自遮挡状态下, 都有一种不可能直接到达的状态. 于是, 式 (3) 可简化为

$$P(s_t|z_t) = \max_{s \in D(t-1)} P(s|s_{t-1})P(z_t|s) \quad (4)$$

其中, 对于任意状态 $x \in O$, 函数 $D(x)$ 将返回 x 能够直接到达的所有状态的集合.

于是, 在时刻 t , 自遮挡状态 s_t 可由下式获得:

$$s_t = \arg \max_{s \in D(t-1)} P(s|s_{t-1})P(z_t|s) \quad (5)$$

2 层次肤色模型

2.1 肤色特征提取

为了提取肤色特征, 本文采用 Hsu^[14] 等提出的基于 YCbCr 空间的加入亮度补偿的椭圆肤色模型. 为了排除背景等噪声的干扰, 以提高椭圆肤色模型的鲁棒性, 在提取肤色特征之前, 我们利用背景差分技术进行背景分割, 并利用数学形态学操作和种子填充对背景分割结果图像进行处理, 提取出目标人体区域, 见图 5(a), 图 5(b) 为相应的彩色图像. 在此基础上, 利用椭圆肤色模型对图像进行肤色检测, 见图 5(c), 并利用数学形态学操作对检测结果进行形态学处理, 见图 5(d), 为了方便后续计算, 按照人体部分进行分割, 并对分割结果进行标准化处理, 按照自顶向下的顺序, 分别记为 I_H, I_A, I_L , 见图 5(e).

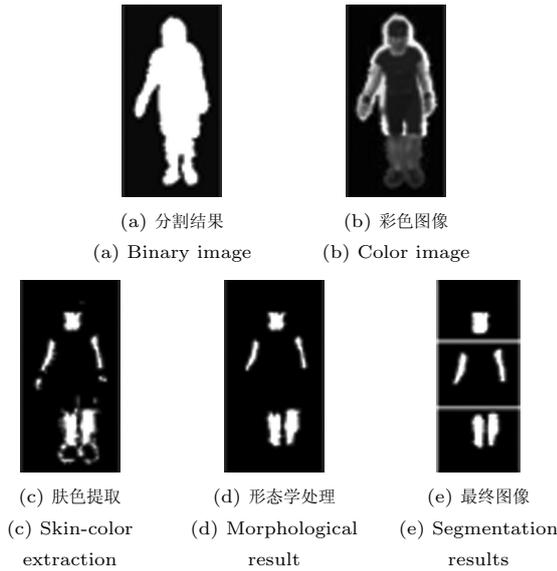


图 5 肤色特征提取过程

Fig. 5 The process of skin color features extraction

2.2 模型的定义

利用上一小节的方法所提取的肤色特征, 结合自遮挡状态概率模型的定义, 按照人体部位自顶向下的顺序, 本文定义层次肤色模型如下:

1) 面部层次概率模型: 为了建立面部层次概率模型, 根据前文定义的人体自遮挡状态, 本文利用主成分分析 (Principal component analysis, PCA) 方法训练不同自遮挡状态

下的面部肤色模型. 由于在人体处于背行状态时, 面部完全被遮挡, 此时, 并无任何面部特征. 因此, 仅需训练其他三个状态的面部肤色样本并建立相应的三个状态模型即可. 为了训练面部肤色特征模型, 我们利用上一小节中描述的特征提取方法, 对全部所选样本提取面部特征, 所提取的面部特征集合记为 S^H ; 并通过手工标注的方法, 对所提取出来的面部样本集合 S^H 进行状态分类: 前行状态面部特征样本集记为 T_1^H , 左侧遮挡状态下的面部特征样本集记为 T_2^H , 右侧遮挡状态下的面部特征样本集记为 T_3^H . 对于任意一个面部样本集 T_i^H , $1 \leq i \leq 3$, 设标准化图像 $I_{i,j}^H \in T_i^H$, 其中, $1 \leq j \leq N_i^H$, N_i^H 为第 i 个面部样本集合所包含的样本个数, 图像 $I_{i,j}^H$ 的大小为 $w \times h$, 则根据 PCA 算法, 面部特征模板可由下述步骤进行训练:

步骤 1. 将 T_i^H 中的任意一个样本 $I_{i,j}^H$ 由原 $w \times h$ 维的图像转换为 $1 \times V$ 维图像, 转换后的图像记为 $\tilde{I}_{i,j}^H$, 其中, $V = w \times h$;

步骤 2. 重复步骤 1, 直至 T_i^H 中的所有样本全部转换完成;

步骤 3. 利用图像拼接技术将转换后的样本图像集合 $\{\tilde{I}_{i,j}^H\}$ 拼接成一个图像矩阵, 记为 I_i^M , I_i^M 的尺寸为 $N_i^H \times V$;

步骤 4. 利用 PCA 算法, 获得图像 I_i^M 的主分量 P_i^C 的均值向量 E_i 和特征矩阵 Q_i , 令 $H_i = \{E_i, Q_i\}$;

步骤 5. 重复步骤 1~4, 直至全部三个样本集合训练完毕后结束本训练过程.

令 $H = \{H_1, H_2, H_3\}$ 表示面部肤色特征模型集合, 按照下标递增的顺序, 分别表示利用 PCA 方法建立的前行状态、左侧遮挡状态和右侧遮挡状态下的面部肤色模型, 则面部层次概率模型可表示为

$$M(I_t^H, i) = \begin{cases} 1 - \frac{\text{sum}(|I_t^H - Bproj(proj(I_t^H, H_i), H_i)|)}{w \times h \times 255}, & 1 \leq i \leq 3 \\ 0, & i = 4 \text{ 且 } \text{sum}(I_t^H) \geq 1 \\ 1, & i = 4 \text{ 且 } \text{sum}(I_t^H) = 0 \end{cases} \quad (6)$$

其中, I_t^H 表示在时刻 t 所提取出来的面部肤色特征图像; i 定义为当前自遮挡状态的序号; $I' = proj(I_t^H, H_i)$ 表示投影函数, 该函数的功能是利用特征模型 H_i 将图像 I_t^H 投影到正交系 V_i 上, 并将结果存储到图像 I' 中; $I'' = Bproj(I', H_i)$ 为逆向投影函数, 该函数的功能是将图像 I_t^H 经过 $proj$ 函数得到的图像 I' 进行逆向投影, 并将结果存储到图像 I'' 中; 分子中的 sum 函数为图像中的像素值求和函数; 分母中的 w 和 h 分别表示图像 I_t^H 的高度和宽度, 分母起到了归一化的作用; 由于背行状态下无法提取面部肤色特征, 因此在背行状态下, 此时, $i = 4$, 若 $\text{sum}(I_t^H) \geq 1$, 即当前图片能够提取到肤色特征, $M(I_t^H, i) = 0$, 否则, $M(I_t^H, i) = 1$; 在其他三个状态下, 此时, $1 \leq i \leq 3$, $M(I_t^H, i)$ 可由投影函数和逆向投影函数计算. 投影函数 $proj(\cdot, \cdot)$ 可定义为

$$I' = proj(I_t^H, H_i) = (\tilde{I}_t^H - E_i) \times Q_i^T \quad (7)$$

其中, \tilde{I}_t^H 为 I_t^H 经过转换后的图像, 尺寸为 $1 \times V$ 维图像矩阵, V_i^T 为第 i 个面部样本模型 H_i 的特征矩阵的转置.

同样, 逆向投影函数 $Bproj(\cdot, \cdot)$ 可定义为

$$I'' = Bproj(I', H_i) = I' \times Q_i + E_i \quad (8)$$

其中, \tilde{I}' 为 I' 经过转换后的图像, 尺寸为 $1 \times V$ 维图像.

2) 手臂层次概率模型: 为了获得手臂皮肤轮廓特征集 B^A 中的各种特征, 我们首先利用第 2.1 节中描述的特征提取方法, 对全部所选样本提取手臂肤色特征, 所提取的手臂肤色样本特征集合记为 S^A ; 其次, 通过手工标注的方法, 对所提取出来的 S^A 进行状态分类: 前行状态手臂肤色样本特征集记为 T_1^A , 左侧遮挡状态下的手臂肤色样本特征集记为 T_2^A , 背行状态手臂肤色样本特征集记为 T_3^A , 右侧遮挡状态下的手臂肤色样本特征集记为 T_4^A . 在此基础上, 对于任意手臂样本特征集 T_i^A , $1 \leq i \leq 4$, 设标准化图像 $I_{i,j}^A$, 其中, $1 \leq j \leq N_i^H$, 于是, 手臂样本特征可按照下述步骤训练:

步骤 1. 根据 Canny 算法, 提取 $I_{i,j}^A$ 中每一个肤色区域的轮廓点集, 记为 C_k , k 为所找到轮廓的个数, 且 $0 \leq k \leq 2$;

步骤 2. 根据步骤 1 所获得的 C_k , 利用基于最小二乘的椭圆最佳拟合算法, 利用拟合后的椭圆型轮廓, 可获得椭圆重心的二维坐标 (x_k, y_k) ;

步骤 3. 根据步骤 1 和步骤 2, 若仅检测到一个手臂轮廓, 则表明一只手臂完全处于自遮挡状态, 即 $C_{C,i,j}^A = 1$, 则 $C_{A,i,j}^A$ 设定为 -1 , $C_{D,i,j}^A$ 设定为 0 ; 若检测到两个手臂轮廓, 即 $C_{C,i,j}^A = 2$, 则 $C_{A,i,j}^A$ 和 $C_{D,i,j}^A$ 为有效计算值, 其计算式为

$$\begin{aligned} C_{A,i,j}^A &= \arctan(|y_1 - y_2|, |x_1 - x_2|) \\ C_{D,i,j}^A &= \sqrt{(y_1 - y_2)^2 + (x_1 - x_2)^2} \end{aligned} \quad (9)$$

其中, \arctan 为反正切函数;

步骤 4. 重复步骤 1~3, 直至当前手臂样本特征集 T_i^A 中的全部样本计算完成, 令

$$B_{D,i}^A = \{C_{C,i,j}^A, C_{A,i,j}^A, C_{D,i,j}^A\}_{j=1}^{N_i}$$

步骤 5. 根据样本集 T_i^A 的特征集 $B_{D,i}^A$ 中的有效值, 即在 $C_{C,i,j}^A = 2$ 的情况下, 分别计算 $C_{A,i,j}^A$ 和 $C_{D,i,j}^A$ 的期望和方差, 分别记为 $\mu_{A,i}^A$, $\mu_{D,i}^A$, $\sigma_{A,i}^A$ 和 $\sigma_{D,i}^A$;

步骤 6. 重复步骤 1~5, 直至四个样本集全部训练完成后结束本训练过程.

根据第 1.2 节中的分析, 本文所定义的手臂层次概率模型包括两部分: 手臂夹角层次概率模型和手臂距离层次概率模型. 定义 $F_A^A(s, c_{A,t}^A, c_{C,t}^A, \mu_{A,\Omega}^A, \sigma_{A,\Omega}^A)$ 为手臂夹角层次概率模型, 定义 $F_D^A(s, c_{D,t}^A, c_{C,t}^A, \mu_{D,\Omega}^A, \sigma_{D,\Omega}^A)$ 为手臂距离层次概率模型, 其中, 概率分布函数可由下式定义:

$$F(s, x, c, \mu_\Omega, \sigma_\Omega) = \begin{cases} N(x; \mu_\Omega, \sigma_\Omega), & c = 2 \\ 1, & c = 1 \end{cases} \quad (10)$$

其中, s 为自遮挡状态; x 为手臂肤色特征值; c 为手臂轮廓个数; μ_i 和 σ_i 表示自遮挡状态的样本数据的期望和方差; $N(x; \mu, \sigma)$ 为满足期望为 μ 、方差为 σ 的高斯函数; 下标 Ω 为状态 s 所对应的序号, $\Omega = E(s)$, $E(\cdot)$ 为序号函数, 返回当前状态 s 所对应的序号.

根据式 (10), 手臂夹角层次概率模型和手臂距离层次概率模型可由下式定义:

$$\begin{aligned} F_A^A(s, c_{A,t}^A, c_{C,t}^A, \mu_{A,\Omega}^A, \sigma_{A,\Omega}^A) &= \begin{cases} N(c_{A,t}^A; \mu_{A,\Omega}^A, \sigma_{A,\Omega}^A), & c = 2 \\ 1, & c = 1 \end{cases} \\ F_D^A(s, c_{D,t}^A, c_{C,t}^A, \mu_{D,\Omega}^A, \sigma_{D,\Omega}^A) &= \begin{cases} N(c_{D,t}^A; \mu_{D,\Omega}^A, \sigma_{D,\Omega}^A), & c = 2 \\ 1, & c = 1 \end{cases} \end{aligned} \quad (11)$$

其中, $\Omega = E(s_t)$.

3) 腿部层次概率模型: 腿部层次概率模型包括腿部夹角层次概率模型和腿部距离层次概率模型. 按照手臂层次概率模型中所描述的方法, 根据式 (10) 和式 (11), 腿部夹角层次概率模型可定义为 $F_A^L(s_t, c_{A,t}^L, c_{C,t}^L, \mu_{A,\Omega}^L, \sigma_{A,\Omega}^L)$, 腿部距离层次概率模型可定义为 $F_D^L(s_t, c_{D,t}^L, c_{C,t}^L, \mu_{D,\Omega}^L, \sigma_{D,\Omega}^L)$.

2.3 模型的计算

根据层次肤色模型中面部肤色模型的分析 and 定义, 本小节给出了自遮挡状态模型的计算过程. 由于在不同的自遮挡状态下, 行人的面部肤色特征具有明显的不同特征, 例如面部肤色的面积, 见图 3(a), 因此, 本文利用面部肤色特征状态来表示式 (4) 中的转移概率 $P(s_t|s)$. 于是, 根据式 (6), 转移概率 $P(s_t|s)$ 可由下式计算:

$$P(s|s_{t-1}) = \frac{M(I_t^H, E(s))}{\sum_{i \in F(D(s_{t-1}))} M(I_t^H, i)} \quad (12)$$

其中, $E(s)$ 为序号函数, 返回当前状态参数的序号. 由函数 $D(x)$, $x \in O$ 的定义, $E(D(s_{t-1}))$ 表示返回由时刻 $t-1$ 的状态能够直接达到的状态的序号集合.

由于 B^A , B^L 之间相互独立, C_A^A , C_D^A 之间相互独立, 同样, C_A^L , C_D^L 之间相互独立, 式 (4) 中的观测概率可进行如下推导:

$$\begin{aligned} P(z_t|s) &= P(B|s) = P(B^A, B^L|s) = \\ &P(B^A|s) \times P(B^L|s) = \\ &P(C_A^A|s) \times P(C_D^A|s) \times P(C_A^L|s) \times P(C_D^L|s) \end{aligned} \quad (13)$$

根据手臂层次概率模型和腿部层次概率模型公式, 式 (13) 中等式右边的 4 个条件概率可由式 (14) 计算, 具体过程见式 (10) 和式 (11).

$$\begin{aligned} P(C_A^A|s) &= F_A^A(s, c_{A,t}^A, c_{C,t}^A, \mu_{A,\Omega}^A, \sigma_{A,\Omega}^A) \\ P(C_D^A|s) &= F_D^A(s, c_{D,t}^A, c_{C,t}^A, \mu_{D,\Omega}^A, \sigma_{D,\Omega}^A) \\ P(C_A^L|s) &= F_A^L(s, c_{A,t}^L, c_{C,t}^L, \mu_{A,\Omega}^L, \sigma_{A,\Omega}^L) \\ P(C_D^L|s) &= F_D^L(s, c_{D,t}^L, c_{C,t}^L, \mu_{D,\Omega}^L, \sigma_{D,\Omega}^L) \end{aligned} \quad (14)$$

其中, $\Omega = E(s)$.

3 实验

本算法的实验视频来源于美国布朗大学 Leonid 等利用 VICON 系统所采集的 HumanEva 人体运动实验数据集^[13]. 算法的编程环境为 Visual Studio .NET 2003 和 OpenCV1.0, 全部实验在一台为配置双核 CPU 1.8GHz, 1G DDR 内存的计算机上完成.

实验视频选取行人室内按圆周行走的视频, 视频图像大小为 640 像素 \times 480 像素, 总计 1005 帧图像, 从中选取 423 帧连续图像作为训练样本, 这 423 帧连续图像构成了步行人体运动过程中完整的一个圆周. 在此基础上, 利用目测估计的方法, 以某一个肢体未被遮挡肤色面积 D_O 大于等于该肢体全部肤色面积 D_A 的 2/3 为标准, 即 $D_O \geq D_A \times 2/3$, 手工标注四种自遮挡状态, 其中, 背行状态的样本数目为 57 帧, 左侧遮挡状态的样本数目为 137 帧, 右侧遮挡状态的样本数目为 173 帧, 前行状态的样本数目为 56 帧. 面部肤色特征图

像标准化后的高度为 $h = 150$, 宽度为 $w = 200$. 右侧遮挡状态的有效样本 ($C_C^A = 2$) 数目为 59 帧, 左侧遮挡状态的有效样本 ($C_C^L = 2$) 数目为 90 帧. 表 1 为各遮挡状态下所选择的训练样本数目在各遮挡状态所包含的全部图像中所占比率.

表 1 各自遮挡状态的训练样本所占比率

Table 1 The percentages of samples of each self-occlusion state in all frames

	总帧数	样本数目	所占比率 (%)
前行状态	115	56	48.7
右侧遮挡状态	288	137	47.6
左侧遮挡状态	484	173	35.7
背行状态	118	57	48.3
总计	1 005	423	42.1

本文利用训练样本所获得的统计数据, 对全部视频图像进行识别, 实验结果如表 2 所示. 从表 2 可以看出, 算法的准确率达到了 94.9%. 产生错误的一个原因是在 HumanEva 数据集中的实验视频中, 由于光照而导致的阴影和图像分辨率较低等原因 (这些情况与实际应用中所拍摄的图像较类似), 因此, 无法利用椭圆肤色模型准确地对前行状态和背行状态下的肤色进行检测, 如图 6 所示. 产生错误的另外一个原因是: 在左侧遮挡状态和右侧遮挡状态下的计算误差, 该误差等于这两个状态下的错误率.

表 2 本文算法检测结果

Table 2 The results of our algorithm detection

	总帧数	错误帧数	错误率 (%)
前行状态	115	5	4.3
右侧遮挡状态	288	7	2.4
左侧遮挡状态	484	28	5.9
背行状态	118	11	9.3
总计	1 005	51	5.1

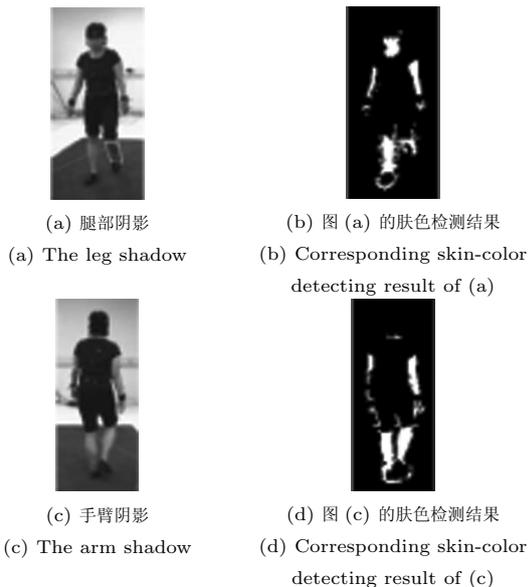


图 6 光照不均匀条件下的椭圆肤色模型检测结果

Fig. 6 The skin-color detection results using ellipse skin model on the poor illumination condition

4 结论

根据马尔可夫模型和层次肤色特征的定义, 结合人体运

动的时序性特征, 本文所提出的算法能够利用单目视频图像对人体运动中四肢自遮挡现象进行检测. 与传统的学习方法相比, 该方法仅对面部肤色特征利用 PCA 方法进行训练, 而对于步行人的手臂和腿部的肤色特征, 则通过统计人体在行进中手臂和腿部的几何特征而获得具体运动参数, 从而缩短了训练样本的时间, 减少了时间开销. 并且, 该方法与多目摄像机系统相比, 减少了硬件成本的开销. 实验表明, 该方法具有很高的准确性.

目前, 由于 HumanEva 的实验数据集视频特殊性, 本文所提出的算法适用于与该数据集的实验视频类似环境, 如人体医疗康复性训练等. 在实际应用中, 经常会出现以下情况: 1) 人体运动过程中皮肤裸露面积较少, 有的视频中甚至仅有面部皮肤和手部皮肤能够检测到; 2) 由于人体的动作的随意性, 可能会出现甩臂等现象, 从而导致状态是别的二义性问题. 针对肤色裸露面积的问题, 可通过在衣服的放置标志物, 或者衣服的不同部位标志不同颜色加以解决, 此时, 可利用不同部位的颜色特征代替人体肤色模型; 对于检测过程中可能出现的二义性问题, 可结合人体动作识别算法获得解决, 如基于上下文的人体动作识别等算法, 另外, 还可考虑将本文的算法结合人体 3D 运动模型对人体动作进行运动约束, 从而对人体剧烈运动情况进行剪枝处理, 以解决二义性问题. 课题的下一步研究工作重点将集中解决上述缺点, 以及将本文算法应用于 HumanEva 数据集的人体 3D 运动跟踪.

References

- 1 Moeslund T B, Granum E. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 2001, **81**(3): 231–268
- 2 Moeslund T B, Hilton A, Kruger V. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006, **104**(2): 90–126
- 3 Agarwal A, Triggs B. 3D human pose from silhouettes by relevance vector regression. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Washington D. C., USA: IEEE, 2004. 882–888
- 4 Grauman K, Shakhnarovich G, Darrell T. Inferring 3D structure with a statistical image-based shape model. In: *Proceedings of IEEE International Conference on Computer Vision*. Washington D. C., USA: IEEE, 2003. 641–648
- 5 Zhuang Yan, Zhan Hong-Bin, Wang Wei, Wang Ke. Weighted color histogram based particle filter for visual target tracking. *Control and Decision*, 2006, **21**(8): 868–872 (庄严, 战洪斌, 王伟, 王珂. 基于加权颜色直方图和粒子滤波的彩色物体跟踪. *控制与决策*, 2006, **21**(8): 868–872)
- 6 Luo Zhong-Xiang. Video Motion Extraction and Motion Synthesis [Ph. D. dissertation], Zhejiang University, China, 2002 (罗忠祥. 视频流中的人体运动提取与运动合成 [博士学位论文], 浙江大学, 中国, 2002)
- 7 Kehl R, Gool L V. Markerless tracking of complex human motions from multiple views. *Computer Vision and Image Understanding*, 2006, **104**(2-3): 190–209
- 8 Lu Y F, Wang L, Hartley R, Li H D, Shen C H. Multi-view human motion capture with an improved deformation skin model. In: *Proceedings of IEEE Conference on the Digital Image Computing: Techniques and Applications*. Washington D. C., USA: IEEE, 2008. 420–427

- 9 Christoudias C M, Urtasun R, Kapoor A, Darrell T. Co-training with noisy perceptual observations. In: Proceedings of the IEEE on International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 1–10
- 10 Urtasun R, Fleet D J, Geiger A, Popovic J, Darrell T J, Lawrence N D. Topologically-constrained latent variable models. In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: IEEE, 2008. 1080–1087
- 11 Urtasun R, Darrell T J. Sparse probabilistic regression for activity-independent human pose inference. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 12 Wang J M, Fleet D J, Hertzmann A. Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(2): 283–298
- 13 Sigal L, Black M J. HumanEva: Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion, Technical Report TR CS-06-08, Brown University, USA, 2006
- 14 Hsu R L, Abdel-Mottaleb M, Jain A K. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 696–706

于雪松 哈尔滨工业大学博士研究生. 主要研究方向为计算机视觉, 人体运动分析. 本文通信作者. E-mail: yxz001@hit.edu.cn
(**YU Xue-Song** Ph. D. candidate at Harbin Institute of Technology. His research interest covers computer vision and human motion analysis. Corresponding author of this paper.)

刘家锋 博士, 哈尔滨工业大学计算机学院副教授. 主要研究方向为人工智能、模式识别、计算机视觉. E-mail: jeffery@public.hr.hl.cn
(**LIU Jia-Feng** Ph. D., associate professor at the School of Computer Science and Technology, Harbin Institute of Technology. His research interest covers artificial intelligence, pattern recognition, and computer vision.)

唐降龙 博士, 哈尔滨工业大学计算机学院教授. 主要研究方向为模式识别、人工智能、中文信息处理、智能人机接口、穿戴式计算机、人体生物特征身份鉴别. E-mail: txl60@public.hr.hl.cn
(**TANG Xiang-Long** Ph. D., professor at the School of Computer Science and Technology, Harbin Institute of Technology. His research interest covers pattern recognition, artificial intelligence, Chinese information processing, intelligent human-computer interface, wearable computer, and human biometric identification.)

黄剑华 博士, 哈尔滨工业大学计算机学院副教授. 主要研究方向为计算机视觉, 模式识别, 掌上型计算机, 人体生物特征身份鉴别. E-mail: jhhuang@hit.edu.cn
(**HUANG Jian-Hua** Ph. D., associate professor at the School of Computer Science and Technology, Harbin Institute of Technology. His research interest covers computer vision, pattern recognition, palmtop, and biometrics.)