

# 一种基于改进 PLSA 和案例推理的行为识别算法

涂宏斌<sup>1</sup> 岳艳艳<sup>2</sup> 周新建<sup>3</sup> 罗 锐<sup>1</sup>

(华东交通大学轨道交通学院 南昌 330013)<sup>1</sup> (华东交通大学国际学院 南昌 330013)<sup>2</sup>

(华东交通大学载运工具与装备教育部重点实验室 南昌 330013)<sup>3</sup>

**摘要** 针对行为人发生的行为因遮挡或者自遮挡可能导致行为歧义性的问题,提出基于改进 PLSA 和案例推理算法的行为识别方法。该算法既可以克服传统 PLSA 算法中生成式模型对观察特征序列的独立性假设会导致过拟合的缺点,又可以消除由于遮挡等原因引起的歧义性带来的识别精度降低问题。实验表明该方法能有效地提高人体行为识别准确率。

**关键词** 行为识别,歧义性行为,时空兴趣点,PLSA,案例推理

**中图分类号** T391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.06.050

## Novel Action Recognition via Improved PLSA and CBR

TU Hong-bin<sup>1</sup> YUE Yan-yan<sup>2</sup> ZHOU Xin-jian<sup>3</sup> LUO Kun<sup>1</sup>

(School of Railway Tracks and Transportation, East China Jiaotong University, Nanchang 330013, China)<sup>1</sup>

(International School, East China Jiaotong University, Nanchang 330013, China)<sup>2</sup>

(The Key Laboratory of Vehicle and Equipment, Ministry of Education, East China Jiaotong University, Nanchang 330013, China)<sup>3</sup>

**Abstract** In order to recognize the ambiguous action meaning in the same scene for occlusion, the improved PLSA and CBR algorithm are used to recognize the simple action according to the space-time interest point. The improved algorithm can not only overcome the shortcomings of the over fitting in the traditional PLSA algorithm owe to the generative model independence assumption of observation sequences, but also decrease recognition ambiguity. Experiments based on the proposed method are executed on the public databases such as KTH, Weizmann, UCF sports and the self-building database. The results show that the proposed method has the performance of validity and effectiveness.

**Keywords** Action recognition, Ambiguous behavior, Space-time interest point, PLSA (Probabilistic Latent Semantic Analysis), CBR (Case Based Reason)

## 1 引言

近年来,智能视频监控(Intelligent Video Surveillance)是计算机视觉领域中倍受关注的一个应用领域<sup>[1]</sup>。为了实现智能监控的智能化,监控系统利用计算机视觉理论对视频图像进行处理、分析和理解<sup>[2]</sup>。智能视频监控系统可以对各种场景中不同人的行为进行识别,当人出现异常行为时,系统能够以最快的方式向相关安全机构发出警报并提供有用的监控信息,从而能够更加有效地协助安保人员处理突发安全事件<sup>[3-4]</sup>。

近年来,主题模型(Probabilistic Latent Semantic Analysis, PLSA)由于其不需要监督和人工标注等优点,得到了迅速的应用和发展。在人体行为识别方面,由于多数主题模型是层

次贝叶斯模型,使得其在对各类动作进行建模时更加得心应手。此外,主题模型能够解决过拟合、不需要进行标定等问题。这都使得主题模型在人体行为识别领域中能够迅速地得到应用和发展。

Zhang 等提出了一种结构化 PLSA 算法对人体行为进行识别<sup>[5]</sup>,通过学习行为的表示作为非监督潜在主题,通过用单词码本表示人体行为中局部形状上下文特征来表示人体行为。Peng Zhang 等提出了一种新的基于 LSA 的人体运动跟踪算法<sup>[6]</sup>,该方法采用一种新的“twin-pipeline”训练框架找到运动人体目标的潜在语义主题模型,通过对人体在不同环境下的兴趣点实现对运动人体目标的跟踪,有效地避免了由于累计误差从而降低跟踪精度。Ping Guo 等利用视觉词对人体行为进行建模并用一个视觉局部模型来表示,从而建立一

到稿日期:2016-05-30 返修日期:2016-10-17 本文受华东交通大学基金(2003414090),国家自然科学基金(51268015),江西省自然科学基金项目(20122BAB206027)资助。

涂宏斌(1979—),男,讲师,CCF 会员,主要研究方向为模式识别、深度学习和图像分析,E-mail: tuhongbin310@163.com;岳艳艳(1984—),女,主要研究方向为模式识别、数据挖掘;周新建(1963—),教授,硕士生导师,主要研究方向为模式识别、计算机视觉和数据挖掘;罗 锐(1978—),讲师,主要研究方向为故障诊断、模式识别。

个平移和尺度不变的概率潜在语义分析模型<sup>[7]</sup>。Ruitai Li 等提出了一个用于分类和识别人的异常行为的混合词包模型<sup>[8]</sup>,并提出 h-PLSA 模型对人体行为进行半监督学习。朱旭东等将主题模型和隐马尔科夫链模型相结合来对人体异常行为进行识别<sup>[9]</sup>,有效地克服了两个模型在行为识别中精度和鲁棒性方面的缺陷。谢飞等针对人体发生行为时常有姿态变化多、运动幅度大小不一等情况<sup>[10]</sup>,结合 3D-SIFT 特征和 HOOOF 特征提出了 3DSH 特征法,并利用 TMBP 模型对行为进行识别。本文针对行为发生的行为因为遮挡或者自遮挡在同一场景下同一行为可能会出现不同的识别,即行为出现歧义性的情况导致识别精度下降,提出了基于改进 PLSA 和案例推理算法相结合的行为识别方法。首先利用改进的 PLSA 识别算法先对行为人的行为进行粗识别;然后利用案例推理原理消除由于遮挡等原因引起的行为歧义性;最后利用人体行为数据库 KTH<sup>[12]</sup>、Weizmann 数据库实验<sup>[13-14]</sup>、UCF sports 数据库<sup>[15]</sup>和我们的数据库进行实验,验证了本文方法的正确性。

## 2 人体运动特征表示

运动特征是指从动作低层数据中提取的某些可以表征人体行为的信息。目前,在人体行为识别领域常用的特征有:剪影特征、光流特征、时空特征、轨迹特征和时空兴趣点特征等。时空兴趣点特征是视频中空间和时间上梯度变化显著的像素点,该特征将人体动作信息以一些不关联的点的形式进行描述,所以只需要提取出少量的兴趣点即可对人体行为进行分析和识别。本文采用时空兴趣点表示人体运动特征。

目前,时空兴趣点特征广泛应用于人体行为识别领域,如 Zhu Yu 等利用深度图像的时空兴趣点(STIP)特征进行人体动作识别<sup>[16]</sup>,Marin-Jimenez 等利用密度采样 STIP 特征对人体交互行为进行识别<sup>[17-18]</sup>。

对于空间域图像  $f^p$ ,给定观测尺度  $\delta_x^2$  和平滑尺度  $\delta_y^2 = \delta_x^2$ ,根据哈里斯角点检测方法通过找到的图像兴趣点可以得到:

$$\mu^p = g^p(\cdot; \delta_x^2) * \begin{pmatrix} (L_x^p)^2 & (L_x^p L_y^p) \\ (L_x^p L_y^p) & (L_y^p)^2 \end{pmatrix} \quad (1)$$

其中,  $L_x^p$  和  $L_y^p$  为如下定义的高斯微分函数。

$$L_x^p(\cdot; \delta_x^2) = \partial_x(g^p(\cdot; \delta_x^2) * f^p) \quad (2)$$

$$L_y^p(\cdot; \delta_x^2) = \partial_y(g^p(\cdot; \delta_x^2) * f^p) \quad (3)$$

其中,  $g^p(\cdot; \delta_x^2)$  和  $g^p(\cdot; \delta_y^2)$  分别为平滑窗函数和观测窗函数。

$$g^p(x, y; \delta^2) = \frac{1}{2\pi\delta^2} \exp\left(-\frac{(x^2 + y^2)}{2\delta^2}\right) \quad (4)$$

$$\mu^p = \det(\mu^p) - k \text{trac}^2(\mu^p) = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \quad (5)$$

当视频图像中的像素点出现图像强度值在空间域和时间域上都有显著变化时(时空兴趣点),定义函数:  $f: R^2 \times R \rightarrow R$ ,并将其与高斯核函数做卷积,得到的结果即为尺度空间  $L$ ,该空间定义为:  $L: R^2 \times R \times R_+ \rightarrow R$

$$L(\cdot; \delta_x^2, \tau_i^2) = g(\cdot; \delta_x^2, \tau_i^2) * f(\cdot) \quad (6)$$

该函数的最大值即为时空兴趣点:

$$H = \det(\mu) - k \text{trac}^2(\mu) = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3 \quad (7)$$

本文利用文献[20]使用 HOG(Histogram of Oriented Gradients)描述每个特征。

## 3 PLSA 模型的基本原理

Hofmann<sup>[21-23]</sup>中详细阐述了一个新的分类模型,即概率潜在语义分析模型(Probabilistic Latent Semantic Analysis, PLSA)。该模型不仅可用于文本分类,还可以分析自然语言,是挖掘对象潜在联系的一个非常实用的方法。该算法一般是将高维的向量通过某种变换有效地映射到一个低维空间内,从而挖掘出对象间许多其他的联系。该算法的基本原理简介如下。假设  $D = \{d_1, d_2, \dots, d_N\}$  表示某个文档集,而  $W = \{w_1, w_2, \dots, w_M\}$  表示一个单词集,将  $D$  和  $W$  合并为一个  $N \times M$  的共现矩阵  $N$ ,该矩阵表示  $w_j$  出现在  $d_i$  中的次数,如式(8)所示:

$$N = (n(d_i, w_j))_{ij} = \begin{pmatrix} n(d_1, w_1) & \dots & n(d_1, w_M) \\ n(d_2, w_1) & \dots & n(d_2, w_M) \\ \vdots & & \vdots \\ n(d_N, w_1) & \dots & n(d_N, w_M) \end{pmatrix} \quad (8)$$

其中,  $d_i$  为第  $i$  个文档,  $w_j$  表示第  $j$  个单词,在矩阵  $N$  中,行表示每个单词出现在某个文档中的次数,列表示某个单词出现在每一个文档中的次数。定义一个主题变量  $z$ ,  $p(d_i)$  为一个单词出现在  $d_i$  篇文档中的概率;  $p(w_j | z_k)$  为在主题  $z_k$  中出现  $w_j$  的概率;  $p(z_k | d_i)$  为主题  $z_k$  出现在第  $i$  个文档中的概率。则三者关系如图 1 所示。



图1 “文档集合-潜在语义-单词集合”关系图

由于 PLSA 模型是在一个合理的统计基础上应用了退火似然函数,并用迭代 EM 算法表示有效的拟合过程,因此该算法优于其他标准语义分析模型。目前,PLSA 模型已成功应用于信息过滤、文本分类、信息检索和行为识别等方面。

## 4 基于改进 PLSA 歧义行为识别算法

本节针对行为发生的行为在同一场景下可能会有不同的含义,例如不同朝向的运动人体受到遮挡或深度影响而容易产生歧义不能有效识别的问题,提出了一种利用改进的 PLSA 识别算法先对行为人的行为进行识别,然后利用案例推理原理消除由于遮挡等原因引起的歧义性,从而得出正确的识别结果。

首先,将每一帧中的人体行为转换为相对应的词包信息,将获得的人体各个关键点的时空兴趣点转换为词包<sup>[23-24]</sup>。本节先利用改进的 k-均值聚类算法建立单词本,该算法可以有效地提高聚类中心初始化的鲁棒性和实时性;然后利用文献[25]中的算法进行识别,算法示意图如图 2 所示。

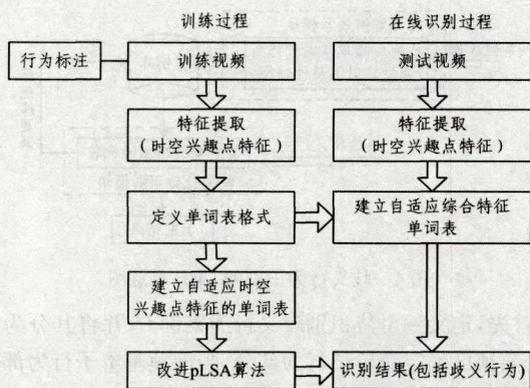


图 2 利用改进 pLSA 算法对行为进行识别的算法示意图

人体简单行为类别  $d$  在视频图像  $p$  中可以由聚类得到  $N_v$  个视觉词,表示为一个  $N_v$  维的向量  $d(p)$ :

$$d(p) = \{n(p, v_1), \dots, n(p, v_i), \dots, n(p, v_{N_v})\} \quad (9)$$

其中,  $n(p, v_i)$  表示人体行为  $d$  的图像  $p$  中含有视觉词  $v_i$  的个数。

通常情况下,采用 K 均值聚类的方法对“词包”进行聚类。本节使用 BOW 模型表示上文提取到的相空间特征,然后连接两类特征视觉词所对应的直方图,视觉词分布所对应的直方图即可表示出相空间特征,如图 3 所示。

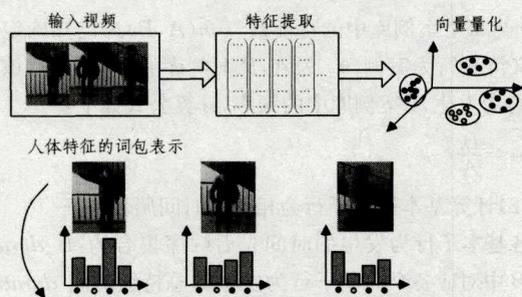


图 3 用“词包”模型表示人体相空间特征的处理过程

综上所述,本文用“词包”进行描述的具体流程如下:

- 1) 根据提取的时空特征点特征描述表格,提取其中的文本词;
- 2) 根据提取出的时空兴趣点特征组成的综合特征,用“词包”方法来获得其视觉词;
- 3) 将第一步得到的文本词和第二步得到的视觉词进行结合,得到综合特征的完整“词包”描述。

假设时空兴趣点特征集合为:  $X = \{x_1, \dots, x_N\}, x_N = \hat{x}_k^N(t), N=1, \dots, n$ , 假设  $Y = \{y_1, \dots, y_N\}, y_i \in \mathcal{Q}^d, i=1, \dots, n$  为集合  $X$  的降维集合,聚类类别为  $\{C_i\}_{i=1}^M$ 。

通过 k-means 聚类算法创建出单词表<sup>[26]</sup>:

$$\min \sum_{i=1}^N \sum_{j=1}^K r_{ij} \|x_i^{\eta}(t) - \mu_j\|^2 \quad (10)$$

其中,  $k$  为聚类数,  $r_{ij} \in \{0, 1\}$  为标记数,如果  $y_i$  为  $j$  类,则  $r_{ij} = 1$ , 否则  $r_{ij} = 0$ 。

$$\min J = \sum_{j=1}^K \sum_{i=1}^N S_{ij} \|x_i^{\eta}(t) - \mu_j\|^2 \quad (11)$$

此处  $K$  为单词表大小,  $S_{ij}$  为  $x_i^{\eta}(t)$  降维后的权值。定义一个簇内方差<sup>[27]</sup>:

$$S_{\max} = \max_{1 \leq j \leq k} v_k = \max_{1 \leq j \leq k} \left\{ \sum_{i=1}^N S_{ij} \|x_i - \mu_j\|^2 \right\} \quad (12)$$

定义簇内方差的权值  $\epsilon_w$ :

$$\epsilon_w = \sum_{i=1}^K \sum_{j=1}^K \omega_k^p S_{ij} \|x_i - \mu_j\|^2 \quad (13)$$

$$\omega_k \geq 0, \sum_{k=1}^M \omega_k = 1, 0 \leq p < 1 \quad (14)$$

其中,指数  $p$  表示类间权值更新的敏感程度。根据重复性实验,得出经验值为 0.65,可得式(15):

$$\min_{\{C_i\}_{i=1}^k} \max_{\{\omega_i\}_{i=1}^K} \epsilon_w \quad (15)$$

$$\omega_k \geq 0, \sum_{k=1}^M \omega_k = 1, 0 \leq p < 1$$

先固定权值找到新的聚类,然后再计算出  $m_k$ 。

$$\sigma_{ij} = 1, k = \operatorname{argmin}_{1 \leq i \leq k} (\omega_k^p \|x_i - \mu_i\|^2) \quad (16)$$

当  $\omega_k$  递增时,  $m_k$  的计算公式如下:

$$\mu_i = \frac{\sum_{j=1}^k \sigma_{ij} x_j}{\sum_{j=1}^k \sigma_{ij}} \quad (17)$$

$$\omega_k = \frac{v_k^{\frac{1}{1-p}}}{\sum_{k=1}^M v_k^{\frac{1}{1-p}}} \quad (18)$$

其中

$$v_k = \sum_{i=1}^N \varphi_{ik} \|x_i - \mu_i\|^2, 0 \leq p < 1, \frac{1}{(1-p)} > 0$$

为了增强改进 k-means 算法的稳定性<sup>[27]</sup>,更新权值定义为:

$$\omega_k^{(t)} = \beta \omega_k^{(t-1)} + (1-\beta) (\omega_k^{1/(1-p)} / \sum_{j=1}^N \omega_j^{1/(1-p)}), 0 \leq \beta \leq 1 \quad (19)$$

其中,在不断的迭代计算过程中,参数  $\beta$  能够影响到控制权重的更新,并能对连续迭代的权重值进行平滑。给出权重矩阵  $\omega_k$ , 可以得式(20):

$$\min_{GG^T} J = \min_{GG^T=1} \operatorname{tr}(W^{\frac{1}{2}} C^T X^T X C W^{\frac{1}{2}}) \quad (20)$$

根据文献[27],具体算法过程如下。

输入: 视频序列集合  $X = \{x_i\}_{i=1}^N$ , 初始化聚类中心  $\{m_k\}_{k=1}^K$  (类别数为  $K$ )

输出: 聚类结果  $\{\delta_{ik}\}, i=1, \dots, N; k=1, \dots, K$ , 聚类中心  $\{m_k\}_{k=1}^K$

Set  $t=0$

Set  $p_{\text{init}}=0$

Set  $\omega_k^{(0)} = \frac{1}{K}, \forall k=1, \dots, K$

Set  $p=p_{\text{init}}$

Repeat

$t=t+1$

For  $i=1$  to  $N$  // 迭代更新聚类结果

For  $k=1$  to  $K$

$\sigma_{ij} = 1, k = \operatorname{argmin}_{1 \leq i \leq k} (\omega_k^p \|x_i - \mu_i\|^2)$ , 否则  $\sigma_{ij} = 0$

End for

End for

If  $p < p_{\text{init}}$  then

$\delta_{ik}^{(t)} = [\Delta^{(p)}]_{ik}, \forall k=1, \dots, K, \forall i=1, \dots, N$

$\omega_k^{(t-1)} = [\omega^{(p)}]_k, \forall k=1, \dots, K$

End if

For all  $m_k, k=1$  to  $K$

$$m_k^{(1)} = \frac{\sum_{i=1}^K \delta_{ik}^{(1)} x_i}{\sum_{i=1}^K \delta_{ik}^{(1)}}$$

End for

If  $p < p_{\max}$

$$\Delta^{(p)} = [\delta_{ik}^{(p)}]$$

$$w^{(p)} = [w_k^{(p-1)}]$$

Until  $|\xi_w^{(p)} - \xi_w^{(p-1)}| < \epsilon$

Return  $\{\delta_{ik}^{(1)}\}_{i=1, \dots, M; k=1, \dots, K}, m_k^{(1)}$

End

当成功构建单词本后,本节采用改进的PLSA算法<sup>[28-29]</sup>

对人体行为进行识别分类,具体过程如下。

先定义一个联合分布模型和后验概率公式:

$$p(w, d, z) = p(w|z, z_0) p(z, z_0|d) p(d) \quad (21)$$

$$\lambda = \sum_w \sum_d n(w, d) \log p(w, d) \quad (22)$$

为了求出EM是期望最大计算方法,同时也是一个通用算法,E步骤中的估计值为:

$$E = \sum_w \sum_d n(w, d) (z + z_0) \quad (23)$$

$$z = \sum_z p(z|w, d) \log [p(w|z) p(z|d) p(d)] \quad (24)$$

$$z_0 = \sum_{z_0} p(z_0|w, d) \log [p(w|z_0) p(z_0|d) p(d)] \quad (25)$$

通过设置合适的拉格朗日乘数 $\alpha$ 和 $\beta$ 对式(25)中的参数进行正则化,可以得到:

$$H = E + \sum_z \alpha [1 - \sum_w p(w|z)] + \sum_d \beta [1 - (\sum_z p(z|d) + \sum_{z_0} p(z_0|d))] \quad (26)$$

$$p(z|d) = \frac{\sum_w n(w, d) p(w, d)}{\sum_z \sum_w n(w, d) p(z|w, d) + \sum_{z_0} \sum_w n(w, d) p(z_0|w, d)} \quad (27)$$

$$p(z_0|d) = \frac{\sum_w n(w, d) p(w, d)}{\sum_z \sum_w n(w, d) p(z|w, d) + \sum_{z_0} \sum_w n(w, d) p(z_0|w, d)} \quad (28)$$

$$p(w|z) = \frac{\sum_d n(w, d) p(z|w, d)}{\sum_d \sum_w n(w, d) p(z|w, d)} \quad (29)$$

然后利用贝叶斯估计  $p(z|w, d)$  和  $p(z_0|w, d)$ , 具体计算步骤如下:

$$p(z|w, d) = \frac{p(w, d, z)}{p(w, d)} = \frac{p(w|z) p(z|d)}{\sum_z p(w|z) p(z|d) + \sum_{z_0} p(w|z_0) p(z_0|d)} \quad (30)$$

$$p(z_0|w, d) = \frac{p(w, d, z_0)}{p(w, d)} = \frac{p(w|z_0) p(z_0|d)}{\sum_z p(w|z) p(z|d) + \sum_{z_0} p(w|z_0) p(z_0|d)} \quad (31)$$

在利用改进PLSA算法识别出了行为人的行为后,由于遮挡、自遮挡等原因,会导致最终识别结果具有一定的歧义性。在文献[30]的基础上建立复杂行为案例库识别因遮挡导致的歧义行为,即先对先前帧中已经识别出的行为类别进行学习,建立案例库,然后利用案例推理原理对其进行推理,从而得出准确的识别结果,具体流程如图4所示。

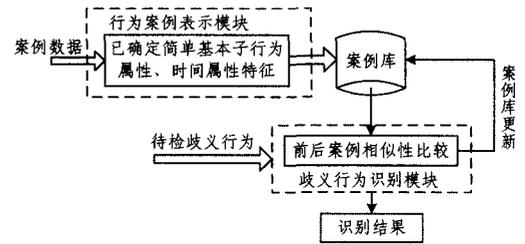


图4 歧义行为识别的CRB示意图

首先,定义一个待识别歧义行为案例C,并将其分为3部分:先前帧已确定简单子行为部分、当前帧基本子行为部分和基本相关子行为时间特征部分。

$$\begin{aligned} \text{Case} = & \{ \{A_{\text{prior}}\} \{B_{\text{current}}\}, \{B_{\text{related}}\} \} \\ = & \{ \{A_1, A_2, \dots, A_n\}, \{B_1, B_2, \dots, B_n\}, \{t_1, t_2, \dots, \\ & t_n\} \} \end{aligned} \quad (32)$$

通过先前帧中确定的行为类别,运用案例推理机制判断出歧义行为,具体步骤如下。

(1) 计算基本简单子行为相似性属性特征

设基本简单子行为组成的一维向量为  $\{A_{\text{prior}}\}$ , 某歧义行为基本子行为部分为  $\{B_{\text{current}}\}$ , 则可计算当前行为和先前帧中已确定行为的相似性:

$$\text{sim}(A, B) = \sum_{i=1}^n w_i \times \text{sim}(A_i, B_i) \quad (33)$$

$n$  为每个案例库中的特征数,  $\text{sim}(A, B)$  为已知确定行为和歧义待测行为案例,  $w_i$  为  $A, B$  案例库的关联权值, 该权值为  $B$  行为发生  $A$  案例库中的频率, 计算公式如下:

$$w_i = \frac{B_i}{A} \quad (34)$$

(2) 计算基本简单子行为相关的时间属性特征

各基本子行为发生的时间先后顺序集合为:  $B_{\text{duration}}$ , 向量  $B$  中对应各个简单子行为发生的总持续时间  $\text{duration} = \{T_1, T_2, \dots, T_N\}$ 。其中  $T_i$  为向量  $B_{\text{duration}}$  中对应的第  $i$  个子行为的总持续时间。

具体算法描述如下:

(1) 对采集到的视频序列计算出人体各个简单行为的两个属性特征;

(2) 计算综合相似性  $\text{sim}(A, B)$ ;

(3) 当相似度最大值大于某一个阈值  $T$  时, 即  $\text{sim}(A, B) > T$ , 则判定  $B$  行为同属  $A$  行为, 并将其加入案例库对案例库进行实时更新, 用于后续歧义行为。

## 5 实验与结果

本节利用现有的人体行为数据库(KTH数据库、WEIZMANN数据库实验、UCF sports数据库和我们的自建数据库)对上述算法进行实验验证。首先利用现有数据库图片进行实验以验证算法的优越性, 然后采用我们自己建立的数据库图片来验证算法的准确性和通用性。

### 5.1 KTH数据库实验

该数据库包含6种人体行为, 即步行、慢跑、快速奔跑、拳击、挥手和鼓掌; 行为在4种不同场景中完成, 即静止的背景、视频比例变化情况下的背景、穿风衣室外环境和室内环境以及光照变化下的背景, 示意图如图5所示。

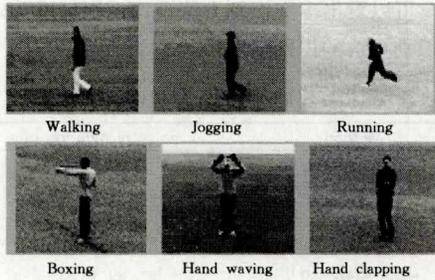


图 5 KTH 数据库中实验视频图例

利用 KTH 数据库进行行为识别,结果如表 1 所列。其中 a1 为步行行为,a2 为慢跑行为,a3 为快速奔跑行为,a4 为拳击行为,a5 为挥手行为,a6 为鼓掌行为。

表 1 利用 KTH 数据库进行识别的结果

a1	0.93	0.10	0.02	0.12	0.13	0.11
a2	0.02	0.92	0.00	0.00	0.01	0.00
a3	0.11	0.00	0.92	0.00	0.00	0.10
a4	0.00	0.00	0.00	0.91	0.10	0.00
a5	0.20	0.10	0.01	0.00	0.89	0.02
a6	0.00	0.10	0.00	0.00	0.01	0.90
	a1	a2	a3	a4	a5	a6

5.2 WEIZMANN 数据库实验

WEIZMANN 数据库由以色列 WEIZMANN 科学院(简称 WIS)计算机科学与应用数学系(简称 CSAM)计算机视觉实验室创建,该数据库分为两个部分:基于事件分析的 WEIZMANN 数据库和时空形状数据库。所有的视频都是通过固定摄像机在简单的背景下拍摄得到的,而且不存在任何遮挡和视角变换的问题,视频中人体的尺寸和动作的快慢可以认为是 不变的。实验图例如图 6 所示。



图 6 WEIZMANN 数据库中实验视频图

利用 WEIZMANN 数据库进行行为识别,结果如表 2 所列。其中 a1 为弯腰行为,a2 为举起双手跳跃行为,a3 为侧身跳跃行为,a4 为双脚起跳行为,a5 为跑步行为,a6 为侧身行走行为,a7 为男子步行行为,a8 为挥手行为,a9 为挥动双手行为,a10 为女子步行行为。

表 2 利用 WEIZMANN 数据库进行识别的结果

a1	0.92	0.10	0.02	0.11	0.01	0.00	0.10	0.10	0.02	0.00
a2	0.00	0.90	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.01
a3	0.00	0.00	0.91	0.13	0.00	0.00	0.12	0.00	0.00	0.00
a4	0.00	0.00	0.00	0.92	0.00	0.01	0.00	0.00	0.00	0.00
a5	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.01	0.01	0.00
a6	0.00	0.00	0.01	0.00	0.00	0.92	0.00	0.00	0.00	0.00
a7	0.10	0.01	0.00	0.00	0.00	0.00	0.91	0.00	0.01	0.00
a8	0.00	0.01	0.12	0.10	0.00	0.00	0.00	0.93	0.00	0.12
a9	0.00	0.10	0.25	0.00	0.10	0.00	0.00	0.00	0.89	0.00
a10	0.10	0.00	0.00	0.01	0.00	0.10	0.00	0.00	0.13	0.90
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10

5.3 UCF sports 数据库

该数据集由美国 UCF(University of Central Florida)大学电子工程与计算机系研究人员创建。其中 UCF sports 创建于 2008 年,数据集总共包含了近 200 个行为视频。UCF 具体介绍见文献[15],本节采用 UCF sports 数据库做实验验证,图例如图 7 所示。

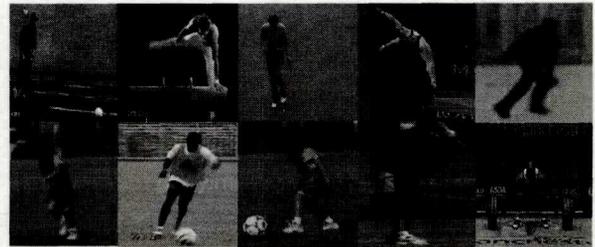


图 7 UCF 数据库中实验视频图例

利用 UCF 数据库进行行为识别,结果如表 3 所列。其中 a1 为男子踢球行为,a2 为跳马行为,a3 为打高尔夫球的前进行为,a4 为高台跳水行为,a5 为侧身奔跑行为,a6 为小孩奔跑行为,a7 为男运动员踢球行为,a8 为小孩踢球行为,a9 为打高尔夫球行为,a10 为举重行为。

表 3 利用 UCF sports 数据库进行识别的结果

a1	0.92	0.10	0.03	0.00	0.10	0.00	0.00	0.12	0.00	0.01
a2	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00
a3	0.11	0.00	0.91	0.12	0.10	0.00	0.00	0.13	0.00	0.04
a4	0.00	0.10	0.00	0.90	0.00	0.00	0.10	0.04	0.00	0.00
a5	0.01	0.01	0.15	0.00	0.88	0.00	0.00	0.13	0.00	0.00
a6	0.10	0.00	0.02	0.00	0.00	0.90	0.00	0.11	0.12	0.10
a7	0.10	0.04	0.00	0.00	0.00	0.00	0.91	0.00	0.02	0.00
a8	0.00	0.13	0.00	0.10	0.00	0.10	0.00	0.92	0.00	0.00
a9	0.00	0.01	0.01	0.12	0.11	0.00	0.00	0.00	0.92	0.12
a10	0.00	0.02	0.00	0.00	0.00	0.30	0.00	0.00	0.00	0.91
	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10

5.4 自建数据库

为了验证本节算法的有效性,我们在不同场景环境下自拍了 30 段视频,每段视频时间为 30~60s,分辨率为 1048 × 1448 像素,帧数最大 25 帧/秒,包括拳击、挥手、投篮、跳跃、小跑和舞蹈等动作。每段视频包括 5 个以上的动作,每段视频的所有动作由一个人或多人连续完成,示意图如图 8 所示。



图 8 自建数据库中实验视频图例

利用自建数据库进行行为识别,结果如表 4 所列。其中 a1 为拳击行为,a2 为挥手行为,a3 为下蹲行为,a4 为侧身快速奔跑行为,a5 为投篮行为,a6 为跳跃行为,a7 为下蹲捡东西行为,a8 为侧身跳跃行为。表 4 列出了利用自建数据库进行识别的结果,从中可以看出,本文提出的识别方法对带有歧义

的行为进行识别的识别精度比文中现有方法的识别精度更高,识别效果好。

表4 利用自建数据库进行识别的结果

a1	0.90	0.13	0.00	0.10	0.12	0.00	0.00	0.01
a2	0.00	0.92	0.00	0.00	0.00	0.01	0.00	0.00
a3	0.01	0.00	0.91	0.10	0.10	0.00	0.13	0.02
a4	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00
a5	0.00	0.11	0.12	0.00	0.90	0.01	0.00	0.12
a6	0.01	0.00	0.03	0.00	0.05	0.92	0.00	0.01
a7	0.03	0.32	0.20	0.00	0.00	0.00	0.91	0.10
a8	0.00	0.10	0.00	0.10	0.00	0.00	0.00	0.92
a1	a2	a3	a4	a5	a6	a7	a8	

## 5.5 算法比较

本节仍然利用上文中的4个数据库,采用本文算法与其他文献中的算法:Tam V. Nguyen<sup>[31]</sup>, Pau Agusti V<sup>[32]</sup>, Alexandros Iosifidis<sup>[33]</sup>, Rami Alazrai<sup>[34]</sup>, Soumitra Samanta<sup>[35]</sup>, Fahad Shahbaz Khan<sup>[36]</sup>和 M. J. Marin-Jiménez<sup>[37]</sup>进行比较,并将本文的识别算法对光流特征和 STIP 特征以及将光流和 STIP 两种特征相结合的特征进行识别,验证之前提出的相空间特征和时空兴趣点特征相结合的综合特征对人体行为进行识别的算法正确性。本节识别方法与其他识别方法的比较结果如表5—表8所列。

表5 利用 KTH 数据库进行算法比较的结果/%

Method	average recognition rate
The proposed method	92.71
Tam V. Nguyen <sup>[31]</sup>	89.83
Pau Agusti V <sup>[32]</sup>	90.20
Alexandros <sup>[33]</sup>	89.88
Rami Alazrai <sup>[34]</sup>	88.82
Soumitra Samanta <sup>[35]</sup>	90.18
Fahad Shahbaz Khan <sup>[36]</sup>	90.21
Marin-Jiménez <sup>[37]</sup>	89.31

从表5可以看出本文识别方法平均识别精度为92.71%,优于其他识别算法。

表6 利用 WEIZMANN 数据库进行算法比较的结果/%

Method	average recognition rate
The proposed method	92.10
Tam V. Nguyen <sup>[31]</sup>	87.80
Pau Agusti V <sup>[32]</sup>	88.10
Alexandros <sup>[33]</sup>	87.61
Rami Alazrai <sup>[34]</sup>	90.13
Soumitra Samanta <sup>[35]</sup>	90.27
Fahad Shahbaz Khan <sup>[36]</sup>	89.71
Marin-Jiménez <sup>[37]</sup>	89.90

从表6可以看出本文识别方法平均识别精度为92.10%,优于其他识别算法的识别精度。

表7 利用 UCF 数据库对算法进行比较的结果/%

Method	average recognition rate
The proposed method	92.13
Tam V. Nguyen <sup>[31]</sup>	89.80
Pau Agusti V <sup>[32]</sup>	90.30
Alexandros <sup>[33]</sup>	89.93
Rami Alazrai <sup>[34]</sup>	90.18
Soumitra Samanta <sup>[35]</sup>	90.15
Fahad Shahbaz Khan <sup>[36]</sup>	89.73
Marin-Jiménez <sup>[37]</sup>	89.88

从表7可以看出本文识别方法平均识别精度为92.13%,高于其他识别算法的识别精度。

表8 利用自建数据库进行算法比较的结果/%

Method	average recognition rate
The proposed method	92.41
Tam V. Nguyen <sup>[31]</sup>	90.20
Pau Agusti V <sup>[32]</sup>	91.36
Alexandros <sup>[33]</sup>	90.12
Rami Alazrai <sup>[34]</sup>	89.13
Soumitra Samanta <sup>[35]</sup>	89.18
Fahad Shahbaz Khan <sup>[36]</sup>	90.26
Marin-Jiménez <sup>[37]</sup>	90.17

从表8可以看出本文识别方法平均识别精度为92.41%,高于其他识别算法的识别精度。

因此,由表5—表8可知,本文的改进 PLSA 识别算法对歧义性的行为识别精度均优于文中比较的现有方法,实验验证了本文算法的正确性和有效性。

**结束语** 本文提出了基于改进 PLSA 和案例推理算法的简单行为识别方法。首先利用改进的 PLSA 识别算法对行为人的行为进行识别,该改进方法可以克服传统 PLSA 算法中生成式模型对观察特征序列的独立性假设会导致过拟合的缺点;然后利用案例推理原理消除由于遮挡等原因引起的歧义性。文中将每帧图像中提取到的人体各个节点时空兴趣点特征转换为时空兴趣点特征词包,采用改进的 k-均值聚类算法建立单词本,利用 WEIZMANN、KTH、UCF sports 和自建数据库对算法进行验证,并将其与其他识别算法和其他特征进行比较,从而验证算法的正确性。实验表明改进后的识别算法准确率高、实时性强,并具有一定的鲁棒性。

## 参考文献

- [1] 张洪博. 视频中的人体行为识别若干关键技术研究[D]. 厦门: 厦门大学, 2013.
- [2] 何卫华. 人体行为识别关键技术研究[D]. 重庆: 重庆大学, 2012.
- [3] SHIAN-RU K, HOANG L U T, YONG-JIN L, et al. A Review on Video-Based Human Activity Recognition[J]. Computers, 2013, 2: 88-131.
- [4] AGGARWAL K J, RYOO S M. Human activity analysis: A review[J]. ACM Computing Surveys, 2011, 43(3): 194-218.
- [5] ZHANG J, GONG S. Action categorization by structural probabilistic latent semantic analysis[J]. Computer Vision and Image Understanding, 2010, 114(8): 857-864.
- [6] ZHANG P, ZHANG Y N, THOMAS T, et al. Moving people tracking with detection by latent semantic analysis for visual surveillance applications[J]. Multimedia Tools and Applications, 2014, 68(3): 991-1021.
- [7] GUO P, MIAO Z, SHEN Y, et al. Continuous human action recognition in real time[J]. Multimedia Tools and Applications, 2012, 68(3): 827-844.
- [8] Li R T, ZHANG C F, ZHONG G H. A novel method of abnormal behaviors recognition[J]. Applied informatics and Communication, PT I, Communications in Computer and Information

- Science, 2011, 224: 79-86.
- [9] ZHU X D, LIU Z J. Human Abnormal Behavior Recognition Based on Topic Hidden Markov Model[J]. Computer Science, 2012, 39(3): 251-255. (in Chinese)  
朱旭东, 刘志镜. 基于主题隐马尔科夫模型的人体异常行为识别[J]. 计算机科学, 2012, 39(3): 251-255.
- [10] 谢飞. 基于主题模型的人物行为识别[D]. 苏州: 苏州大学, 2014.
- [11] LAPTEV I, CAPUTO B. Recognition of human actions[OL]. November 2011. <http://www.nada.kth.se/cvap/actions>.
- [12] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing Human Actions: A Local SVM Approach[J]. Proceedings of International Conference on Pattern Recognition, Icp, 2004, 3: 32-36.
- [13] GORELICK L, BLANK M, SHECHTMAN E, et al. Weizmann actions as space-time shapes [OL]. November 2011. <http://www.wisdom.weizmann.ac.il/vision/SpaceTimeActions.html>.
- [14] ZELNIK-MANOR L, IRANI M. Weizmann event-based analysis of video [OL]. November 2011. <http://www.wisdom.weizmann.ac.il/vision/VideoAnalysis/Demos/EventDetection/EventDetection.html>.
- [15] University of Central Florida. UCF sports action dataset[OL]. February 2012. <http://vision.eecs.ucf.edu/datasetsActions.html>.
- [16] ZHU Y, Chen W, Guo G. Evaluating spatiotemporal interest point features for depth-based action recognition[J]. Image & Vision Computing, 2014, 32(8): 453-464.
- [17] MARÍN-JIMÉNEZ M J, YEGUAS E, BLANCA N P D L. Exploring STIP-based models for recognizing human interactions in TV videos[J]. Pattern Recognition Letters, 2013, 34(15): 1819-1828.
- [18] XIANG T, GONG S, BREGONZIO M. Recognising action as clouds of space-time interest points[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2009: 1948-1955.
- [19] HARRIS C, STEPHENS M. A combined corner and edge detector[J]. Proc of Fourth Alvey Vision Conference, 1988, 1988(3): 147-151.
- [20] GOPARVAR-FARD M, HEYDARIAN A, NIEBLES J C. Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers[J]. Advanced Engineering Informatics, 2013, 27(4): 652-663.
- [21] HOFMANN T. Probabilistic latent semantic indexing[J]. International Acm Sigir Conference on Research and Development in Information Retrieval, 1999, 42(1): 50-57.
- [22] HOFMANN T. Learning the similarity of documents: an information-geometric approach to document retrieval and categorization[M]// Leen T K, Dietterich T G, Tresp V. eds., Advances in Neural Information Processing Systems, vol. 12. The MIT Press, 2000: 914-920.
- [23] HOFMANN T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine Learning, 2001, 42(1/2): 177-196.
- [24] NIEBLES J C, WANG H, FEI-FEI L. 1 Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words[J]. In Proc. BMVC, 2008, 79(3): 299-318.
- [25] FERNANDEZ-BELTRAN R, PLA F. Incremental probabilistic Latent Semantic Analysis for Video Retrieval[J]. Image & Vision Computing, 2015, 38: 1-12.
- [26] KONG Y, ZHANG X, HU W, et al. Adaptive learning codebook for action recognition[J]. Pattern Recognition Letters, 2011, 32(8): 1178-1186.
- [27] GRIGORIOS T, ARISTIDIS L. The MinMax k-means clustering algorithm[J]. Pattern Recognition, 2014, 47(7): 2505-2516.
- [28] DRIESEN J, VAN HAMME H. Modelling vocabulary acquisition, adaptation and generalization in infants using adaptive bayesian PLSA[J]. Neurocomputing, 2011, 74(11): 1874-1882.
- [29] BASSOU N, KOTROPOULOS C. RPLSA: A novel updating scheme for Probabilistic Latent Semantic Analysis[J]. Computer Speech & Language, 2011, 25(4): 741-760.
- [30] XIA L M, YANG B J, CAI N P. Recognition of Suspicious Activity Using Case-based Reasoning in Video Surveillance[J]. Journal of Chinese Computer Systems, 2014, 35(8): 1891-1896. (in Chinese)  
夏利民, 杨宝娟, 蔡南平. 监控视频中基于案例推理的人体可疑行为识别[J]. 小型微型计算机系统, 2014, 35(8): 1891-1896.
- [31] NGUYEN T V, SONG Z, YAN S. STAP: Spatial-Temporal Attention-Aware Pooling for Action Recognition[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2015, 25(1): 77-86.
- [32] PAU AGUSTÍ V, TRAVER J, PLA F. Bag-of-words with aggregated temporal pair-wise word co-occurrence for human action recognition[J]. Pattern Recognition Letters, 2014, 49(C): 224-230.
- [33] IOSIFIDIS A, TEFAS A, PITAS I. Discriminant Bag of Words based representation for human action recognition[J]. Pattern Recognition Letters, 2014, 49: 185-192.
- [34] ALAZRAI R, MOWAFI Y, LEE C S G. Anatomical-plane-based representation for human-human interactions analysis[J]. Pattern Recognition, 2015, 48(8): 2346-2363.
- [35] SAMANTA S, CHANDA B. Space-Time Facet Model for Human Activity Classification[J]. IEEE Transactions on Multimedia, 2014, 16(6): 1525-1535.
- [36] FAHAD SHAHBAZ K, JOOST V D W, MUHAMMAD A R, et al. Semantic pyramids for gender and action recognition[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2014, 23(8): 3633-3645.
- [37] MARÍN-JIMÉNEZ M J, MUÑOZ-SALINAS R, YEGUAS-BOLLIVAR E, et al. Human interaction categorization by using audiovisual cues[J]. Machine Vision & Applications, 2014, 25(1): 71-84.