

文章编号 1004-924X(2018)01-0230-08

## 结合限制密集轨迹与时空共生特征的行为识别

李庆辉\*, 李艾华, 崔智高, 姜 柯  
(火箭军工程大学 五系, 陕西 西安 710025)

**摘要:**针对传统密集轨迹方法应用到真实场景后过多无效轨迹耗费存储与计算资源且严重影响有效特征提取的不足,提出一种新的人体行为识别算法。首先,检测视频帧中存在的人体目标并对获得的包含人体的矩形框进行扩展,利用扩展后的矩形框对传统密集采样特征点的范围进行筛选限制;然后,对筛选限制后的特征点在光流场中跟踪一定帧数获取限制密集轨迹,并在以限制密集轨迹为中心的时空体内构建一组包含轨迹的空间位置、时空上下文信息的特征描述子;最后在视觉词袋模型框架下,采用 SVM 对特征向量进行编码分类。结果显示:在 KTH、YouTube 和 HMDB51 3 个行为数据库上的识别准确率分别达到 98.1%、89.7% 和 66.9%。证明本算法对复杂真实场景中的人体行为具有较高的识别能力。

**关键词:**行为识别;限制密集轨迹;时空共生特征;视觉词袋模型

**中图分类号:**TP391 **文献标识码:**A **doi:**10.3788/OPE.20182601.0230

### Action recognition via restricted dense trajectories and spatio-temporal co-occurrence feature

LI Qing-hui\*, LI Ai-hua, CUI Zhi-gao, JIANG Ke

(Fifth Department, Rocket Force of Engineering University, Xi'an 710025, China)

\* Corresponding author, E-mail: lqhui1212@126.com

**Abstract:** To overcome the limitation of improved dense trajectories for using in real environment, a novel human action recognition algorithm using restricted dense trajectories and spatio-temporal co-occurrence descriptors was proposed. Firstly, a human detector was applied to get the rectangular and the traditional dense interest points in the videos were refined via expanded rectangular box, which greatly reduces the number of trajectories while preserves the discriminative power. Then, the restricted dense trajectories were obtained by tracking the refined points using optical flow fields. And a set of new descriptors was built which describe the relative spatial position and the spatio-temporal context of motion trajectories. Finally, a Bag of Visual Words (BoVW) model with support vector machine was used to classify human action. On three action recognition datasets: KTH, YouTube and HMDB51, the recognition accuracy is 98.1%, 89.7% and 66.9% respectively. Experimental results show that the proposed algorithm has higher recognition ability for human action in complex real scenes.

**Key words:** action recognition; restricted dense trajectories; spatio-temporal co-occurrence feature; bag of visual words

收稿日期:2017-08-29;修订日期:2017-10-09.

基金项目:国家自然科学基金资助项目(No. 61501470);陕西省重点研发计划资助项目(No. 2017GY-075)

## 1 引言

人体行为识别(Human Action Recognition, HAR)是指利用行为识别技术,计算机可以对视频序列中存在的人体行为进行自动地描述和理解,从而将底层视频数据与高层语义关联起来。在视频监控、人机交互、运动分析和视频搜索等领域具有极大的应用价值。由于人体行为的类内差异性和类间相似性以及背景因素的干扰,人体行为识别是非常具有挑战性的课题<sup>[1-2]</sup>。

行为识别的关键是如何对行为建立高效的视频表达。在静态图像识别领域,研究发现密集采样得到的局部特征往往优于人工设计的局部兴趣点特征。受此启发,文献[3]在同一框架下对比了3种时空兴趣点和密集采样特征点的行为识别效果,密集采样算法在9个行为数据库上均取得了最优识别结果。在此基础上,文献[4]通过在光流场中对密集采样得到的特征点跟踪一定帧数,进一步提出了更有效的密集轨迹特征。文献[5]引入单应性矩阵来克服摄像机运动的影响,提出了改进密集轨迹特征(Improved Dense Trajectories, IDT)。目前,对于许多真实场景下的行为数据库,基于密集轨迹的算法均取得了最好的识别结果。

然而,密集采样会带来过多无效点,尤其在复杂的真实场景下大量采样点会落在与行为无关的背景区域,过多无效轨迹不但耗费存储与计算资源还严重影响有效特征的提取。同时传统的特征描述子(HOG、HOF、MBH)仅是对单个像素的计算建立直方图<sup>[4]</sup>,难以提取更深层次的信息。针对以上问题,本文提出了一种结合限制密集轨迹(Restricted Dense Trajectories, RDT)与时空共生描述子的人体行为识别算法。一方面,利用扩展后的人体矩形框对密集采样特征点进行筛选提纯,在降低无效特征点数量的同时又不致丢失对识别行为有重要意义的人体周围环境和交互物体的信息;另一方面,构建一组新的包含轨迹的相对空间位置、时空上下文信息的特征描述子。

## 2 限制密集轨迹

### 2.1 人体扩展矩形框

首先利用可变形部件模型(Deformable Part-based Model, DPM)检测每个视频帧中存在的人体目标,得到包含人体目标的矩形框。DPM算法由3个部分组成:根滤波器获取人体目标的全局轮廓特征,如图1(a);部件滤波器检测人体目标的细节特征,分辨率是根滤波器的2倍,如图1(b);空间模型,描述根滤波器和部件滤波器间的关系,如图1(c)。

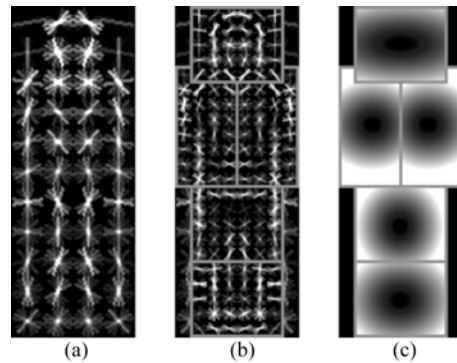


图1 人体 DPM 模型可视化示例

Fig. 1 Visualization of human DPM model

同一种动作在不同的场景下,往往具有不同的含义<sup>[6]</sup>。另外,与人体有交互的物体对行为的识别也有重要的意义。因此为充分利用人体周围场景及交互物体信息,同时排除大量无关信息干扰,在人体检测完成后,对得到的包含人体的矩形框进行扩展操作:宽度扩展为2倍、高度扩展为1.5倍,如图2所示。然后以扩展后的矩形框作为条件,对密集采样得到的特征点进行筛选提纯。

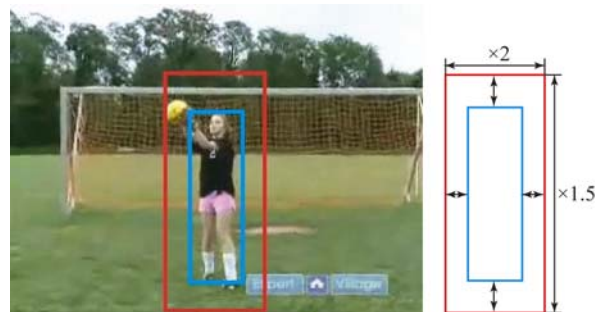


图2 人体扩展矩形框

Fig. 2 Expanded human rectangular box

### 2.2 限制密集轨迹提取

密集轨迹是在时间维度上对密集采样特征点跟踪一定长度得到的<sup>[5]</sup>。本文的 RDT 具体生成步骤如下：

Step1. 密集采样。在每个视频帧的多个空间尺度上,利用以  $K$  个像素点为间隔的网格密集地提取像素点,并过滤掉不含有结构信息的平滑区域采样点。本文的尺度取 8 个,间隔  $K$  取 5,尺度缩放因子为  $1/\sqrt{2}$ ;

Step2. 筛选特征点。以扩展的人体矩形框作为筛选范围,如果密集采样得到的特征点在矩形框内部就保留该特征点,否则删除该点;

Step3. 跟踪特征点,得到限制密集轨迹。计算视频光流场,利用中值滤波器跟踪特征点。

若  $P_t = (x_t, y_t)$  表示某个尺度上第  $t$  帧的特征点,则其在下一帧的位置为:

$$P_{t+1} = (x_{t+1}, y_{t+1}) = (x_t, y_t) + (\mathbf{M} * \boldsymbol{\omega}_t) |_{(x_t, y_t)}, \quad (1)$$

其中,  $\boldsymbol{\omega}_t = (u_t, v_t)$  为两帧之间的光流,  $\mathbf{M}$  是一个  $3 \times 3$  的中值滤波器。为避免跟踪误差的累积,将跟踪的轨迹长度设为  $L = 15 \text{ frame}$ , 并利用光流的方差作为判断依据来剔除静态轨迹和变化剧烈的轨迹。

通过以上步骤实现 RDT 的提取,得到以特征点  $P_t$  为起点的轨迹可表示为  $(P_t, P_{t+1}, \dots, P_{t+15})$ 。图 3 为传统 IDT 与本文 RDT 对比图。可见,本文限制密集轨迹基本保留了与行为相关的轨迹,去除了大部分的背景轨迹。

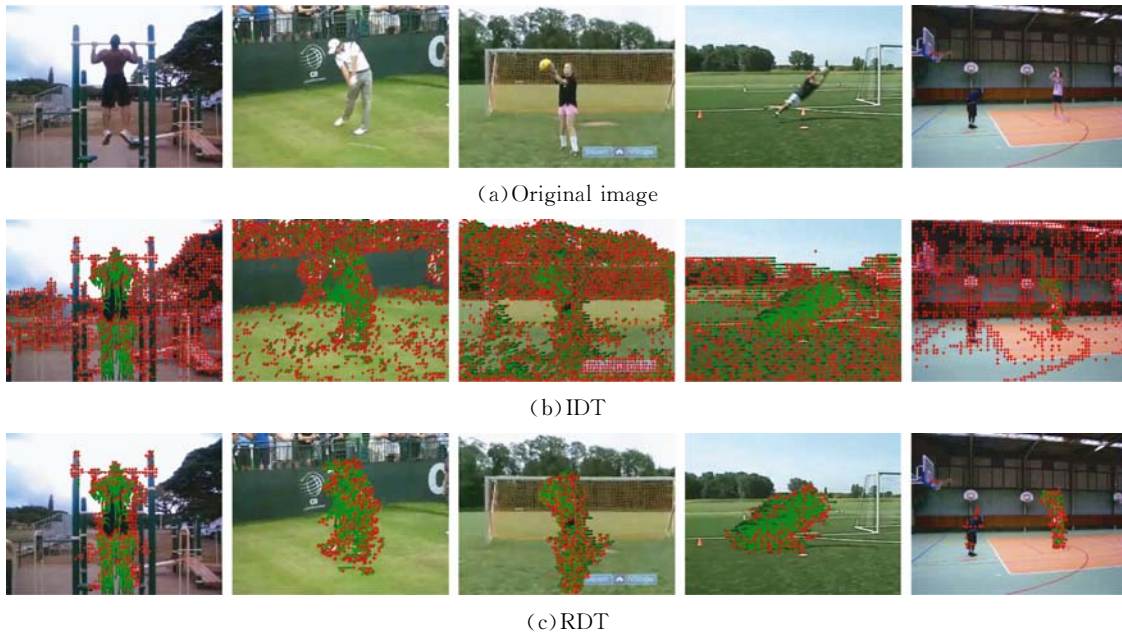


图 3 传统 IDT 与本文 RDT 对比  
Fig. 3 Comparison of original IDT and our RDT

### 3 时空特征描述子

获取限制密集轨迹后,为有效利用视频的运动信息,在每条轨迹周围的三维时空体内构造特征向量。如图 4 所示,以轨迹为中心提取  $W \times H \times L$  的时空体块,并将时空体块分成  $n_s \times n_s \times n_r$  的时空网格,以充分利用其结构信息。最终对每个时空体块统计特征描述子,建立特征直方图。

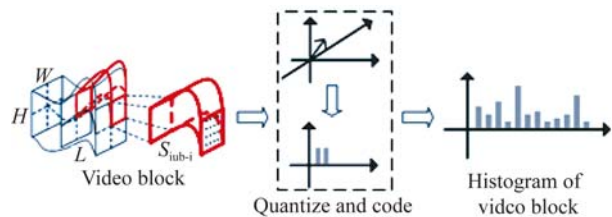


图 4 描述子提取示意图  
Fig. 4 Schematic diagram of descriptor extracting

### 3.1 相对轨迹描述子

BoVW 模型在生成视觉字典时只对局部特征的出现进行统计,而忽视了对行为分类有重要意义的特征间位置关系。常见的解决办法是引入多尺度金字塔,但是这种方法只能提供比较粗糙的信息。本文从特征描述子的构建出发,提出一种融合轨迹空间位置信息的描述子。

在得到人体扩展矩形框后,提取扩展矩形框的几何中心  $P_i^c$  作为动态坐标系的中心。采用跟踪特征点的方法获取几何中心的运动轨迹,表示为  $(P_i^c, P_{i+1}^c, \dots, P_{i+15}^c)$ 。给定特征点轨迹及其对应的扩展矩形框几何中心轨迹,定义相对轨迹(Relative Trajectory)描述子:

$$\mathbf{R} = [(P_i, P_{i+1}, \dots, P_{i+15}) - (P_i^c, P_{i+1}^c, \dots, P_{i+15}^c)]. \quad (2)$$

归一化:

$$\mathbf{R}' = \frac{\mathbf{R}}{\sum_{i=t}^{t+15} \|P_i - P_i^c\|}. \quad (3)$$

归一化后的相对轨迹描述子维数为 30,可以对轨迹的形状信息和相对空间位置信息进行编码,能够区分不同位置的相似轨迹。

### 3.2 空间共生特征描述子

空间 HOG 共生矩阵(SCoHOG)由同一帧两个位置像素梯度方向的联合概率密度来定义。它不仅反映梯度方向的分布特性,也反映具有同样梯度方向的像素之间的位置分布特性,是视频图像的二阶统计特征。SCoHOG 对视频图像上保持固定距离的像素点对的梯度方向进行统计得到共生矩阵,并将展开后共生矩阵拼接作为特征向量。如图 5 所示,本文同时采用  $0^\circ$  扫描和  $90^\circ$  扫描两种方式,距离差分值取  $(2, 0)$  和  $(0, 2)$ 。

另外两种空间共生特征是空间 HOF 共生矩阵(SCoHOF)和空间 MBH 共生矩阵(SCoMBH),这两种特征的计算方法与 SCoHOG 类似。SCoHOF 的统计对象是像素点对的光流方向;SCoMBH 的统计对象是光流水平分量和垂直分量的梯度方向,即 SCoMBH<sub>x</sub> 和 SCoMBH<sub>y</sub>。同样地,采用两种距离差分值  $(2, 0)$  和  $(0, 2)$  来计算这两种空间共生特征。

### 3.3 时间共生特征描述子

时间 HOG 共生矩阵(TCoHOG)是通过在时间维度上不同像素点对的梯度方向统计得到的。由于相邻帧的像素变化不大,本文采取间隔一定

时间的方式提取像素点对,如图 5 所示。TCoHOG 可以有效地描述一定时间内表观的变化。

时间 HOF 共生矩阵(TCoHOF)和时间 MBH 共生矩阵(TCoMBH)与对应的空间共生特征计算方法类似,不同的是在时间维度上提取共生单元。TCoHOF 的共生单元可以反映运动方向的变化,而 TCoMBH 的共生单元反映的是局部运动边界强度的变化。

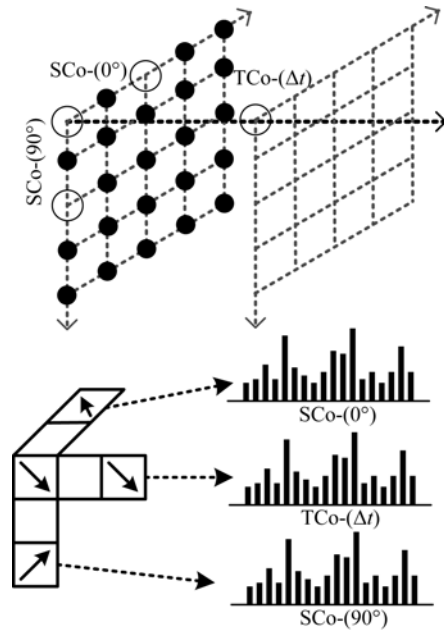


图 5 共生特征描述子提取

Fig. 5 Spatio-temporal co-occurrence descriptor extraction

对于大小为  $m \times n$  的图像块,在距离差分值为  $(a, b)$  时,定义共生矩阵  $\mathbf{C}$ :

$$\mathbf{C}(s, t) = \sum_{i=1}^m \sum_{j=1}^n \begin{cases} \frac{G(i, j) + G(i+a, j+b)}{2}, & \text{if } O(i, j) = s \\ & \text{and } O(i+a, j+b) = t \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

式中:  $G(i, j)$  和  $O(i, j)$  分别为像素点  $(i, j)$  梯度或光流的幅值和方向,  $s, t$  为梯度或光流的量化方向标号。

## 4 实验与分析

### 4.1 人体行为数据库

KTH 数据库<sup>[7]</sup>由 6 类行为视频组成:走

(walking)、慢跑(jogging)、跑(running)、拳击(boxing)、挥手(hand waving)和鼓掌(clapping)。每类行为视频由 25 人分别在 4 种场景下拍摄:室外场景、室外场景尺度变化、室外场景衣物变化和室内场景。数据库总共包含 2 391 段视频,每段视频平均长度 4 s,只包含单人的单种行为。本文采取数据库作者的原始策略,将数据库分成两部分:训练集(16 subjects)和测试集(9 subjects)。

YouTube 数据库<sup>[8]</sup>由在 25 个不同场景下拍摄的 11 类行为视频组成,包括骑车(biking)、投篮(basketball shooting)、跳水(diving)、高尔夫(golf swinging)、骑马(horse riding)、颠球(soccer juggling)、荡秋千(swinging)、网球(tennis swinging)、蹦床(trampoline jumping)、排球(volleyball spiking)、遛狗(walking dog)。包含 1 168 段视频,存在大量的视角、尺度及光照变化。同原始策略一样,本文采取留一交叉验证的实验方案。

HMDB51 数据库<sup>[9]</sup>共有 51 类 6 766 段行为视频组成,主要从数字电影或 YouTube 网站收集。HMDB51 数据库包含着人的面部动作、肢体动作以及交互行为,多数为真实场景下非固定相机拍摄的视频,并且每类行为存在着巨大的类内差异,因此非常具有挑战性。本文将数据库分 3

组分别进行训练-测试。

#### 4.2 视觉词袋模型(Bag of Visual Words)

为验证本文限制密集采样和编码空间信息算法的有效性,采用标准视觉词袋模型分类框架<sup>[10]</sup>进行测试。首先为每种特征描述子构建视觉字典,从训练样本中对每种特征随机抽取 100K 个特征向量,采用 K-means 算法进行聚类生成视觉字典。在此基础上对行为视频量化编码,将特征向量分配到字典中欧氏距离(Euclidean distance)最近的词,统计词汇出现信息,把得到统计直方图归一化后作为视频的描述。

本文采用一种  $\chi^2$  核的非线性 SVM 作为分类器,通过多通道技术融合不同种类的特征。其中,核可表示为:

$$K(x_i, x_j) = \exp\left[-\sum_c \frac{1}{A^c} D(x_i^c, x_j^c)\right], \quad (5)$$

式中:  $A^c$  是所有训练样本在  $c$  特征通道中  $\chi^2$  距离的均值;  $D(x_i^c, x_j^c)$  表示  $c$  特征通道中视频  $x_i$  和  $x_j$  之间的  $\chi^2$  距离,具体定义为:

$$D(x_i, x_j) = \frac{1}{2} \sum_{k=1}^K \frac{(x_{i,k} - x_{j,k})^2}{x_{i,k} + x_{j,k}}, \quad (6)$$

其中  $x_{i,k}$ 、 $x_{j,k}$  为视频  $x_i$  和  $x_j$  的特征直方图的第  $k$  个 bin 值。本文采用 one-against-rest 策略解决多分类问题,以得分最高的类为分类结果。



图 6 行为视频示例(KTH/YouTube/HMDB51)

Fig. 6 Sample frames from action datasets(KTH, YouTube and HMDB51)

### 4.3 实验结果及分析

为探究本文提出的限制密集轨迹(RDT)和共生特征描述子(SCo-/TCo-)分别对识别效果的影响,本文与传统的轨迹提取方法(IDT)和特征描述子(HOG/HOF/MBH)进行了4组交叉组合实验。实验中参数设置为  $W = H = 32, L = 15, n_o = 2, n_r = 3$ ,方向量化为8个bin(其中HOF量化为9个bin)。根据实验观察视觉字典的大小设为4000,得到的实验结果如图7所示。实验结果表明,RDT和SCo-/TCo-的引入均提高了识别准确率,二者组合时识别结果达到最高,在3个数据库的结果分别为:98.1%、89.7%、66.9%。

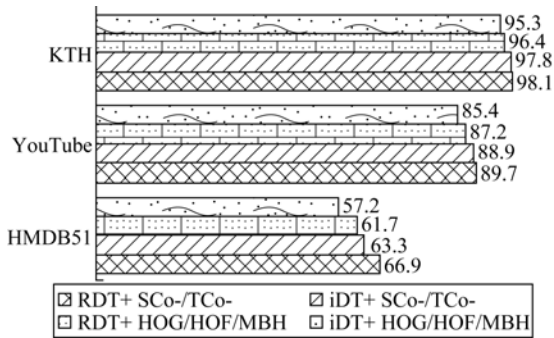


图7 交叉组合实验结果对比

Fig. 7 Experimental results of cross combination

为评估不同种类描述子融合识别效果,本文在BoVW框架下设计了5种融合方案,结果如表1所示。相对轨迹描述子(Relative Trajectory)在3个数据库的识别结果比传统轨迹描述子(Trajectory)均有提高,而且背景环境越复杂提高的幅度越大,在YouTube和HMDB51上分别提高了7.5%和9.3%。这表明了轨迹相对位置信息的有效性。空间共生描述子组合(SCoHOG+

SCoHOF+SCoMBH)和时间共生描述子组合(TCoHOG+TCoHOF+TCoMBH)由于引入了像素对的共生信息,能够提取更具区分力的深层次结构和运动信息,识别结果比原始的HOG+HOF+MBH组合有不同程度地提高。其中,时间共生描述子组合提高的幅度更大,可见时间维度上的运动和结构信息对于作为时间序列的行为视频的分类具有更重要的意义。以上3种描述子融合后,取得的识别结果最优。

表1 不同描述子组合识别结果

Tab. 1 Different combinations of descriptors (%)

Method	KTH	YouTube	HMDB51
Trajectory (Baseline)	90.2	67.2	32.4
HOG+HOF+MBH (Baseline)	93.9	83.4	45.8
RelativeTrajectory	91.9	74.7	41.7
SCoHOG+SCoHOF+SCoMBH	95.1	84.9	54.8
TCoHOG+TCoHOF+TCoMBH	97.4	86.8	59.6
SCo+TCo-	97.8	87.9	62.5
<b>All combined</b>	<b>98.1</b>	<b>89.7</b>	<b>66.9</b>

最后,将本文实验结果与近年来3个数据库上的代表性结果进行了对比,如表2所示。对比算法中既有传统的IDT<sup>[5]</sup>、概率表示<sup>[11]</sup>、非负矩阵表示<sup>[12]</sup>等;也有基于深度学习的算法,如3DCNN<sup>[13]</sup>、LTTN<sup>[14]</sup>等。根据对比结果可知,本文算法在3个数据库上均取得了最优或与最优相近的识别结果,尤其在YouTube、HMDB51数据库上比之前的结果分别提高了1.5%、2.1%。

表2 本文算法结果与当前先进算法结果的对比

Tab. 2 Results comparison of our algorithm and the state-of-the-art (%)

	KTH	YouTube	HMDB51
Le et al. [15]	93.9	Le et al. [15] 75.8	Park et al. [18] 57.2
Ji et al. [13]	90.2	Bhattacharya et al. [11] 76.5	Peng et al. [10] 61.1
Wang et al. [5]	95.3	Wang et al. [5] 85.4	Simonyan et al. [19] 59.4
Wang et al. [16]	95.5	Yang et al. [12] 88.0	Zhu et al. [20] 63.3
LI et al. [17]	<b>98.2</b>	Wang et al. [16] 88.2	Varol et al. [14] 64.8
<b>Our method</b>	<b>98.1</b>	<b>Our method 89.7</b>	<b>Our method 66.9</b>

## 5 结 论

本文提出一种结合限制密集轨迹与时空共生描述子的人体行为识别算法。算法一方面对检测到的包含人体的矩形框进行扩展,利用扩展后的矩形框对传统密集采样特征点的范围进行筛选限制,在降低无效特征点数量的同时又不致丢失对

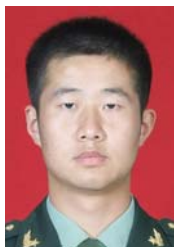
识别行为有重要意义的人体周围环境和交互物体信息;另一方面,构建一组新的包含轨迹的相对空间位置、时空上下文信息的特征描述子。在 3 个人体行为数据库的实验结果表明,本文算法对复杂真实场景中的人体行为具有较高的识别能力,对 YouTube 和 HMDB51 两个数据库分类的准确率分别达到 89.7% 和 66.9%。

### 参考文献:

- [1] HERATH S, HARANDI M, PORIKLI F. Going deeper into action recognition: A survey [J]. *Image and Vision Computing*, 2017, 60: 4-21.
- [2] 刘智, 黄江涛, 冯欣. 构建多尺度深度卷积神经网络行为识别模型 [J]. *光学精密工程*, 2017, 25(3): 799-805.  
LIU ZH, HUANG J T, FENG X. Action recognition model construction based on multi-scale deep convolution neural network [J]. *Opt. Precision Eng.*, 2017, 25(3): 799-805. (in Chinese)
- [3] WANG H, ULLAH M M, KLÄSER A, *et al.*. Evaluation of local spatio-temporal features for action recognition[C]. *Proceedings of British Machine Vision Conference*, BMVC, 2009: 7-10.
- [4] WANG H, KLÄSER A, SCHMID C, *et al.*. Action recognition by dense trajectories[C]. *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2011: 3169-3176.
- [5] WANG H, SCHMID C. Action recognition with improved trajectories [C]. *Proceedings of 2013 IEEE International Conference on Computer Vision*, IEEE, 2013: 3551-3558.
- [6] 王世刚, 鲁奉军, 赵文婷, 等. 应用在线随机森林投票的动作识别 [J]. *光学精密工程*, 2016, 24(8): 2010-2017.  
WANG SH G, LU F J, ZHAO W T, *et al.*. Action recognition based on on-line random forest voting [J]. *Opt. Precision Eng.*, 2016, 24(8): 2010-2017. (in Chinese)
- [7] 张国梁, 贾松敏, 张祥银, 等. 采用自适应变异粒子群优化 SVM 的行为识别 [J]. *光学精密工程*, 2017, 25(6): 1669-1678.  
ZHANG G L, JIA S M, ZHANG X Y, *et al.*. Action recognition based on adaptive mutation particle swarm optimization for SVM [J]. *Opt. Precision Eng.*, 2017, 25(6): 1669-1678. (in Chinese)
- [8] LIU J G, LUO J B, SHAH M. Recognizing realistic actions from videos "in the Wild"[C]. *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009: 1996-2003.
- [9] RODRIGUEZ M, ORRITTE C, MEDRANO C, *et al.*. One-shot learning of Human activity with an MAP adapted GMM and simplex-HMM [J]. *IEEE Transactions on Cybernetics*, 2017, 47(7): 1769-1780.
- [10] PENG X J, WANG L M, WANG X X, *et al.*. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice [J]. *Computer Vision and Image Understanding*, 2016, 150: 109-125.
- [11] BHATTACHARYAS, SUKTHANKAR R, JIN R, *et al.*. A probabilistic representation for efficient large scale visual recognition tasks[C]. *IEEE Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2011: 2593-2600.
- [12] YANG X D, TIAN Y L. Action recognition using super sparse coding vector with spatio-temporal awareness [C]. *Proceedings of 13th European Conference on Computer Vision*, Springer, 2014: 727-741.
- [13] JI SH W, XU W, YANG M, *et al.*. 3D convolutional neural networks for Human action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221-231.
- [14] VAROL G, LAPTEV I, SCHMID C. Long-term temporal convolutions for action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, doi: 10.1109/TPAMI.2017.2712608.
- [15] LE Q V, ZOU W L, YEUNG S Y, *et al.*. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis[C]. *Proceedings of 2011 IEEE Confer-*

- ence on Computer Vision and Pattern Recognition*, IEEE, 2011: 3361-3368.
- [16] WANG H R, YUAN CH F, HU W M, *et al.*. Action recognition using nonnegative action component representation and sparse basis selection [J]. *IEEE Transactions on Image Processing*, 2014, 23(2): 570-581.
- [17] LIU L, SHAO L, LI X L, *et al.*. Learning spatio-temporal representations for action recognition: A genetic programming approach [J]. *IEEE Transactions on Cybernetics*, 2016, 46(1): 158-170.
- [18] PARK E, HAN X F, BERG T L, *et al.*. Combining multiple sources of knowledge in deep CNNs for action recognition [C]. *Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision*, IEEE, 2016: 1-8.
- [19] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos [C]. *Advances in Neural Information Processing Systems 27*, NIPS, 2014: 568-576.
- [20] ZHU W J, HU J, SUN G, *et al.*. A key volume mining deep framework for action recognition [C]. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2016: 1991-1999.

#### 作者简介:



李庆辉(1989—),男,山东临沂人,博士研究生,2011年、2013年于第二炮兵工程大学分别获得学士、硕士学位,主要从事机器视觉及模式识别方面的研究。  
E-mail:lqhui1212@126.com

#### 导师简介:



李艾华(1966—),男,河北沧州人,博士,教授,博士生导师,1987年、1990年于第二炮兵工程学院分别获得学士、硕士学位,1998年于西安交通大学获得博士学位,主要从事机电设备故障诊断、自动控制、机器视觉等方面的研究。  
E-mail:ahli66@126.com