DOI：10.7652/xjtuxb201802004

一种运用显著性检测的行为识别方法

王晓芳[1,2]，齐春[1]
（1.西安交通大学电子与信息工程学院，710049，西安；
2.齐鲁工业大学(山东省科学院)电气工程与自动化学院，250353，济南）

摘要：针对传统稠密轨迹行为识别法不能很好地区分行为区域和背景的问题，提出一种运用显著性检测的行为识别方法。考虑到视频显著性在较小的时空范围内变化不大，将视频在时域分割为多个短子视频，并将子视频在空域划分成小块，再以块为基础运用一种两阶段显著性检测方法获取每个子视频的行为区域。在检测的第一阶段，将低秩矩阵恢复算法应用于子视频的运动信息计算其初始显著性，并据此将其内所有块划分为候选前景集合和绝对背景集合；在第二阶段，为了将真正的行为区域从候选前景集合中分离出来，利用绝对背景集合中块的运动信息构建字典，通过加权稀疏表示算法计算候选前景集合中每个块的细化显著性，再通过阈值化获取二值显著图用以指示行为区域；最后，将显著图融入稠密跟踪过程以获取行为区域轨迹用于行为识别。基准数据集上的实验结果表明，该方法能够较好地检测视频中的行为区域，获得的识别率高于传统稠密轨迹法2.5%～4.5%。

关键词：行为识别；显著性检测；稀疏表示；低秩矩阵恢复
中图分类号：TP391.4 **文献标志码**：A **文章编号**：0253-987X(2018)02-0024-06

An Action Recognition Method Using Saliency Detection

WANG Xiaofang[1,2], QI Chun[1]
(1. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. School of Electrical Engineering and Automation, Qilu University of Technology(Shandong Academy of Sciences), Jinan 250353, China)

Abstract: An action recognition method based on saliency detection is proposed to address the problem that the traditional dense trajectory method for action recognition does not discriminate action-related areas and background. A video is temporally split into several short sub-videos in considering the problem that the video saliency does not change considerably in a small spatial-temporal region, and each short sub-video is further spatially divided into small patches. Then a two-stage saliency detection method is used based on patches to obtain the action-related areas in each sub-video. A low-rank matrix recovery algorithm is applied to the motion information of a sub-video to calculate the initial saliency of the sub-video and then the value of the initial saliency is used to classify all the patches in the sub-video into a candidate foreground set and an absolute background set in the first stage of detection. In the second stage, weighted sparse representation algorithm based on the dictionary constructed by the motion of patches in the absolute background set is used to compute the refined saliency of each patch in the sub-video and to separate the true action-related areas from the candidate foreground set. A binary saliency map is

收稿日期：2017-06-19。 作者简介：王晓芳(1974—)，女，博士生；齐春(通信作者)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(61572395)。
网络出版时间：2017-12-05 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20171205.2108.002.html

generated by thresholding to indicate action-related areas. Finally, the trajectories in action-related areas are extracted for action recognition by incorporating the saliency map into dense tracking. Experimental results on benchmark datasets show that the proposed method detects the action-related areas in videos well and the recognition rate is 2.5%-4.5% higher than that of the dense trajectory.

Keywords: action recognition; saliency detection; sparse representation; low-rank matrix recovery

行为识别即利用计算机自动提取视频中的行为特征并判别行为类别，在视频监控、人机交互、虚拟现实等领域具有广阔的应用前景。稠密轨迹法[1]是近年来一种比较成功的行为识别方法，该方法通过提取视频稠密采样点的轨迹来获取行为的长时段特征。然而，传统的稠密轨迹法在提取轨迹时不能很好地区分行为区域和背景，对包含相机运动的视频，除行为区域之外背景区域也会产生大量的轨迹，这种背景轨迹和感兴趣的行为关系不大，其存在限制了行为识别性能。

为了改进传统的稠密轨迹法，许多文献提出只获取行为区域内的稠密轨迹用于描述行为特征，这类方法目前主要存在 2 种思路。鉴于背景轨迹通常由相机运动产生，第一种思路先通过估计相机运动校正视频的光流，再利用校正后的光流消除背景轨迹[2-3]。考虑到行为区域通常比背景区域显著，另一种思路先通过检测视频显著性获取行为区域，再提取行为区域内轨迹[4-5]，这种思路的关键在于显著性检测。文献[4]将低秩矩阵恢复应用于运动信息检测视频的显著性，但是不能解决行为区域内部运动一致性的问题；文献[5]能够确定视频中的真实显著图，但依赖于观察者的眼部运动数据；文献[6]利用字典学习和稀疏编码获取视频显著性，但是没有充分利用运动信息；此外，现有文献中也存在许多其他显著性检测方法[7-9]，但大多不是面向行为区域获取而设计。获取视频中行为区域的关键在于如何区分行为区域和背景，而不能只考虑一般意义上的显著性。无论视频是在静态或者动态场景中获取，运动信息都是区分行为区域和背景的重要依据。对于包含相机运动的视频，从总体上看，其背景运动的空域分布具有较高的一致性，而行为运动的空域分布具有一定的不规则性，所以行为区域相对于背景通常具有较高的运动显著性，可以通过运动显著性检测方法将其从背景中分离。然而，一些大的行为区域内部也存在局部一致运动，而有些背景区域也包含局部不规则运动，此时一般的运动显著性检测方法难以将它们很好地区分。

鉴于此，本文提出一种采用两阶段显著性检测获取视频中的行为区域的方法，并将其应用于轨迹法行为识别。本文方法主要包括 2 个阶段：第 1 阶段，将低秩矩阵恢复算法[10]应用于运动信息计算子视频内每个块的初始显著性，并借此将子视频所有块划分为候选前景集合和绝对背景集合；第 2 阶段，利用绝对背景集合中所有块的运动向量构建字典，通过稀疏表示算法[11]获取候选前景集合中所有块的细化显著性。在此基础上，对显著性进行阈值化得到二值显著图用于指示行为区域，最后将其融入稠密跟踪过程以提取行为区域轨迹用于行为识别。与其他显著性检测方法相比，上述两阶段方法能够更充分地考虑行为区域和背景区域的运动特点，从而以更高的对比度突出视频中的行为区域。

1　显著性和行为区域检测

设长度为 T 的视频 $\boldsymbol{V}=[\boldsymbol{I}_1,\boldsymbol{I}_2,\cdots,\boldsymbol{I}_T]$，$\boldsymbol{I}_t$ 表示第 t 帧，在时域将 $\boldsymbol{V}$ 分割成长度均为 w 的 K 个互不重叠的子视频，即 $\boldsymbol{V}=[\boldsymbol{V}_1,\boldsymbol{V}_2,\cdots,\boldsymbol{V}_K]$，第 k 个子视频为 $\boldsymbol{V}_k=[\boldsymbol{I}_{(k-1)w+1},\boldsymbol{I}_{(k-1)w+2},\cdots,\boldsymbol{I}_{kw}]$。在空域将每个子视频划分成 $M\times N$ 个大小相等且互不重叠的时空块，划分后的 $\boldsymbol{V}_k$ 可用一个 3D 分块矩阵表示

$$\boldsymbol{V}_k=\begin{bmatrix}\boldsymbol{P}_1 & \boldsymbol{P}_2 & \cdots & \boldsymbol{P}_N\\ \boldsymbol{P}_{N+1} & \boldsymbol{P}_{N+2} & \cdots & \boldsymbol{P}_{2N}\\ \vdots & \vdots & & \vdots\\ \boldsymbol{P}_{(M-1)N+1} & \boldsymbol{P}_{(M-1)N+2} & \cdots & \boldsymbol{P}_{MN}\end{bmatrix} \tag{1}$$

式中：$\boldsymbol{P}_n$ 为第 n 个时空块，大小为 $s\times s\times w$，其中 s 为空域大小，w 为实域长度。下面以 $\boldsymbol{V}_k$ 为例，利用两阶段显著性检测方法获取子视频中的行为区域，其总体流程如图 1 所示。

1.1　初始显著性检测及候选前景和绝对背景划分

本文采用文献[4]中的方法计算子视频的初始显著性。一般来说，由运动相机拍摄的视频，背景运动空域分布具有一致性，相关性较强，可以认为处于一个低秩的子空间，行为运动空运分布具有随意性，

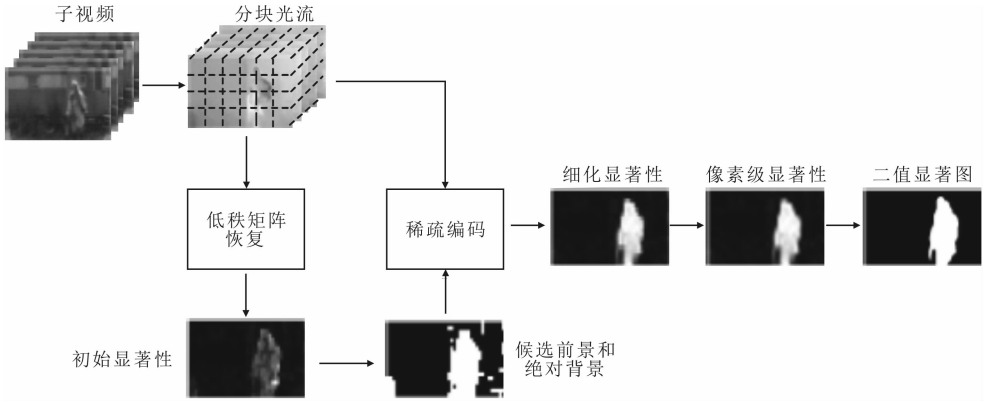


图 1 本文行为区域检测流程

相关性较弱，可以看作稀疏误差。基于上述特点，通过低秩矩阵恢复算法将子视频的运动信息分解成低秩部分和稀疏误差部分，利用后者计算视频块的初始显著性，并据此划分子视频的候选前景和绝对背景。

为了检测 $\boldsymbol{V}_k$ 的初始显著性，需构建其运动矩阵。为此，先获取每个块的运动向量，以块 $\boldsymbol{P}_n$ 为例，先将其每一帧内所有像素点的光流按照空域位置顺序排列得到对应帧的运动向量，其内第 l 帧的运动向量为

$$\boldsymbol{x}_l^{(n)} = [u_{1,l}^{(n)}, v_{1,l}^{(n)}, \cdots, u_{i,l}^{(n)}, v_{i,l}^{(n)}, \cdots, u_{s\times s,l}^{(n)}, v_{s\times s,l}^{(n)}] \tag{2}$$

式中：$u_{i,l}^{(n)}$ 和 $v_{i,l}^{(n)}$ 分别为第 n 块 $\boldsymbol{P}_n$ 的第 l 帧内第 i 个点光流的水平和竖直分量。将块内所有帧的运动向量按时域顺序进行级联，再经转置可得块 $\boldsymbol{P}_n$ 的运动向量 $\boldsymbol{x}_n=[\boldsymbol{x}_1^{(n)}, \boldsymbol{x}_2^{(n)}, \cdots, \boldsymbol{x}_w^{(n)}]^{\mathrm{T}}$，其长度为 $2s^2w$。$\boldsymbol{V}_k$ 内所有块的运动向量获取后，根据各个块的空域位置顺序将其按列堆叠得到 $\boldsymbol{V}_k$ 的运动矩阵

$$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{MN}] \tag{3}$$

通过求解如下低秩矩阵恢复优化问题，可将 $\boldsymbol{X}$ 分解为一个低秩矩阵 $\boldsymbol{B}$ 和一个稀疏矩阵 $\boldsymbol{F}$

$$\begin{aligned} &\arg\min_{\boldsymbol{B},\boldsymbol{F}} \|\boldsymbol{B}\|_* + \lambda\|\boldsymbol{F}\|_1 \\ &\text{s.t.}\quad \boldsymbol{X} = \boldsymbol{B} + \boldsymbol{F} \end{aligned} \tag{4}$$

式中：λ 是用于平衡低秩和稀疏的参数，其值设置为 $\lambda=1.1/[\max(2s^2w, MN)]^{1/2}$。式(4)优化问题可通过增广拉格朗日乘子法(ALM)[12]求解。

矩阵 $\boldsymbol{X}$ 包含子视频 $\boldsymbol{V}_k$ 的总体运动，分解后，背景运动主要包含在低秩矩阵 $\boldsymbol{B}$ 中，行为运动主要包含在稀疏矩阵 $\boldsymbol{F}$ 中。$\boldsymbol{F}$ 中的每一列的值反映了子视频一个块包含行为运动的多少，这里采用列的 l_∞ 范数衡量块的初始显著性。设 $\boldsymbol{f}_n$ 为 $\boldsymbol{F}$ 的第 n 列，则块 $\boldsymbol{P}_n$ 的初始显著性为 $\hat{s}_n = \|\boldsymbol{f}_n\|_\infty$。

按照这种方法，行为区域块因包含行为运动可以获得较高的显著性值，而背景块因不包含行为运动获得较低的显著性值。然而，对于一些大的行为区域，其内部某些行为运动因具有局部一致性被沉积到低秩矩阵 $\boldsymbol{B}$ 中，而对于一些背景区域，其内部运动因具有局部不规则性而被包含到稀疏矩阵 $\boldsymbol{F}$ 中，由此导致行为区域和背景的显著性差异较小，所以利用初始显著性很难将所有行为区域和背景很好地分离。这里通过选定一个较小的阈值 T_s，将所有可能行为区域块(显著性大等于 T_s)都划分到一个候选前景集合 S_f 中，而将剩余绝对背景块(显著性小于 T_s)划分到一个绝对背景集合 S_b 中。

1.2 候选前景显著性细化

利用初始显著性进行集合划分时，由于 T_s 较小，一些背景块也被划分到集合 S_f 中。为了将 S_f 中真正的行为区域块分离出来，需要计算其中的每一个块的细化显著性，以增加行为区域和背景的显著性对比度。一般情况下，对于 S_f 中真正的行为区域块，其运动信息即使和邻近块具有相似性，但都明显不同于绝对背景块；对于 S_f 中的背景块，其运动信息即使含有一定的变化，也和绝对背景块具有较高的相似性。基于此，本节利用 S_b 中所有块的运动向量构建字典，对 S_f 中每一个块的运动向量进行稀疏表示，再利用重构误差计算块的细化显著性。这样，行为区域块因为较难重构而容易获得较高的显著性值；相反，背景块因较易重构而容易获得较低的显著性值。

为了计算 S_f 中每一个块的细化显著性，将 S_b 中所有块的运动向量按列堆叠，得到 $\boldsymbol{V}_k$ 的绝对背景运动矩阵 $\boldsymbol{X}_b$，再将 $\boldsymbol{X}_b$ 作为字典，对 S_f 中的每个块的运动向量进行稀疏表示。以 S_f 中第 r 个块为例，可通过求解以下的优化问题得到其运动向量 $\boldsymbol{x}_{fr}$ 的

稀疏表示

$$\min_{\boldsymbol{\alpha}_r} \| \boldsymbol{x}_{fr} - \boldsymbol{X}_b \boldsymbol{\alpha}_r \|_2^2 + \lambda \| \boldsymbol{\alpha}_r \|_1 \quad (5)$$

式中：$\boldsymbol{\alpha}_r$ 为稀疏表示系数向量。

考虑到背景块一般与它的邻近背景块相关性更强，为了使 S_f 中的背景块获得更低的重构误差，利用 S_b 中的每个块和当前被重构块的空域距离作为 $\boldsymbol{X}_b$ 中对应原子的权重。$\boldsymbol{X}_b$ 中第 i 个原子 $\boldsymbol{x}_{bi}$ 的权重为

$$w_i = \exp\left(\frac{\mathrm{dist}(c_r, c_i)}{\sigma}\right) \quad (6)$$

式中：c_r 和 c_i 分别为当前被重构块和 S_b 中第 i 个块的中心；$\mathrm{dist}(c_r, c_i)$ 为 c_r，c_i 之间的归一化欧式距离；σ 为调节参数。$\boldsymbol{X}_b$ 中所有原子的权重组成一个权重向量 $\boldsymbol{w}_r$，将其引入式(5)，可以得到加权稀疏表示的目标函数

$$\min_{\boldsymbol{\alpha}_r} \| \boldsymbol{x}_{fr} - \boldsymbol{X}_b \boldsymbol{\alpha}_r \|_2^2 + \lambda \| \mathrm{diag}(\boldsymbol{w}_r) \boldsymbol{\alpha}_r \|_1 \quad (7)$$

利用文献[13]中的优化工具箱可以求解式(7)获得稀疏表示系数向量 $\boldsymbol{\alpha}_r$，由此计算重构误差 s_r，将其作为当前被重构块（S_f 中第 r 个块）的细化显著性

$$s_r = \| \boldsymbol{x}_{fr} - \boldsymbol{X}_b \boldsymbol{\alpha}_r \|_2 \quad (8)$$

重复上述过程，可以获取候选前景集合 S_f 中所有块的细化显著性，将其和绝对背景集合 S_b 中所有块的初始显著性按照块的空域位置进行组合，可以得到子视频 $\boldsymbol{V}_k$ 的显著性矩阵 $\boldsymbol{S}_k$。$\boldsymbol{S}_k$ 是一个块级的显著性矩阵，利用空域插值法将其调整为视频帧的原始大小，即获得 $\boldsymbol{V}_k$ 的像素级显著性矩阵，再进行阈值化可以得到 $\boldsymbol{V}_k$ 的二值显著图 $\boldsymbol{M}_k$。$\boldsymbol{M}_k$ 中位置为 (x, y) 的元素 m_{xy} 用于指示子视频 $\boldsymbol{V}_k$ 任意一帧内的点 (x, y) 是否属于行为区域，如果 $m_{xy}=1$，属于行为区域，否则属于背景。

按照上述两阶段法可以计算视频中所有子视频的二值显著图，从而获取视频行为区域。

2　轨迹提取和行为识别

和文献[5]类似，将检测得到的二值显著图和稠密跟踪相结合来提取行为区域轨迹。具体来说，在稠密采样点跟踪过程中，先通过光流获取下一帧上的候选轨迹点，再利用二值显著图判断其是否处于行为区域，如果是则认为是有效轨迹点，否则判其无效并终止当前轨迹。计算识别率时，对每一条轨迹计算 4 种特征形状（Shape）、梯度方向直方图（HOG）、光流方向直方图（HOF）和运动边界直方图（MBH），并利用 FV（Fisher vector）对每一种特征进行独立编码以获取视频级行为特征，最后将 4 种视频级行为特征输入多核学习支撑向量机（SVM）判别行为类别。

3　实验结果和分析

3.1　实验设置

为了验证本文方法的有效性，在 Hollywood2[14] 和 YouTube[15] 2 个实际场景视频数据集上进行实验测试。Hollywood2 共包含 1 707 个视频，分为 12 个行为类别；YouTube 共包含 1 168 个视频，分为 11 个行为类别，每个类别的视频又分为 25 组。检测显著性和行为区域时，设置子视频长度为 5 帧，块的空域大小为 5×5 像素，第 1、第 2 阶段的显著性阈值分别设置为 10 和 50。提取行为区域轨迹时，设置空域采样间隔为 5 像素。计算行为识别率时，对于 Hollywood2 数据集，将其中 823 个视频用作训练样本，剩余 884 个视频用作测试样本；对于 YouTube 数据集，每次利用一组作为测试样本，其余各组用作训练样本，最终识别率是 25 组识别率的均值。

3.2　显著性和行为区域检测结果

采用本文方法对 2 个数据集中 5 个行为视频投篮、骑马、走出汽车、奔跑和站起的行为区域进行检测，各阶段的检测结果如图 2 所示。除最后一个外，其余视频都包含了不同程度、不同类型的相机运动。由图 2 可以看出：第 1 阶段检测到的初始显著性整体对比度较低，尤其是行为区域的中间部分，由于运动存在局部一致性，导致其显著性值更小；第 2 阶段得到的细化显著性能够突出大部分行为区域（包括中间部分），较好地抑制了背景区域。以上结果表明，本文两阶段检测方法能够充分考虑行为区域和

图 2　采用本文方法进行行为区域检测的各阶段结果

背景区域的运动的特点，无论视频是否包含相机运动，都能获得较好的检测结果。

为了进一步验证本文行为区域检测方法的优越性，图 3 将本文检测结果和现有文献最新方法进行对比。其中，文献[8]是一种基于超像素图和时空生长的一般视频显著性检测方法，文献[16]采用一种基于加权稀疏表示的显著性检测方法获取视频中的行为区域。由图 3 可以看出：本文方法检测的显著性具有较高的对比度，能够明显地区分行为区域和背景区域；文献[8]方法的显著性虽然也能够突出视频中的行为区域，但其对比度低于本文方法；文献[16]方法的显著性在行为区域内部较低。

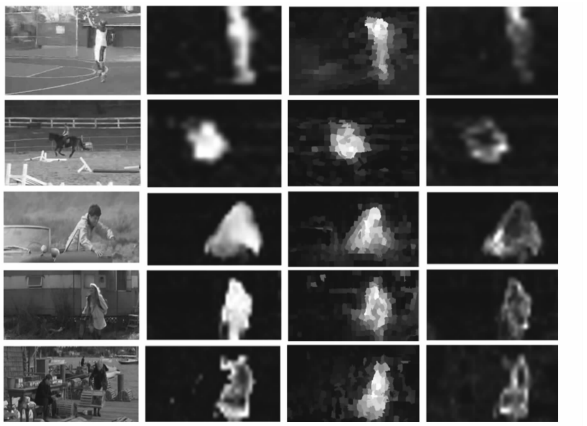


图 3 本文方法和文献[8,16]方法的检测结果对比

3.3 行为区域轨迹提取结果

采用本文方法和传统稠密跟踪方法对 5 个视频的行为区域轨迹进行检测，结果如图 4 所示。由图 4 可以看出：本文方法提取的轨迹不仅具有较好的连续性，而且绝大部分位于行为区域；当视频中存在相机运动时，传统的稠密跟踪方法不仅在行为区域，而且在背景区域也会产生大量轨迹。

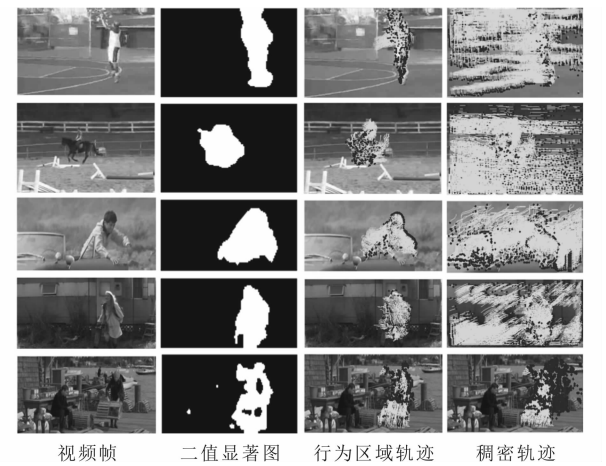


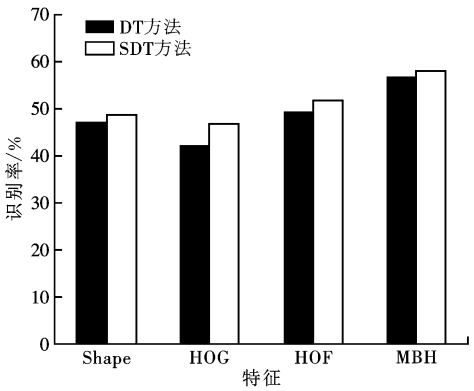
图 4 本文方法和传统稠密方法的行为区域轨迹比较

3.4 行为识别结果

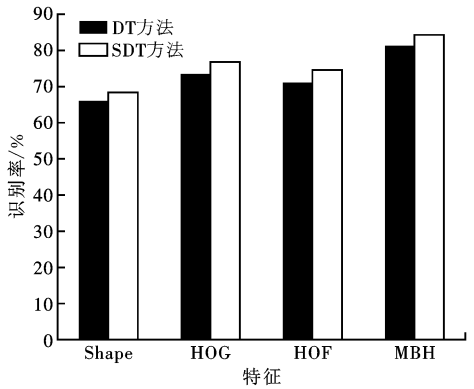
为了验证本文方法的识别性能，分别在 2 个数据集 Hollywood2 和 YouTuber 计算本文方法(SDT)、传统稠密轨迹方法(DT)以及两者视频级特征级联方法(SDT+DT)的总体识别结果，如表 1 所示。由表 1 可见，在 2 个数据集上，SDT 的识别结果都明显优于 DT，而二者级联能够进一步提高识别率。图 5 比较了本文方法(SDT)和传统稠密轨迹跟踪方法(DT)对 2 个数据集上的 4 个特征的识别结果。由图 5 不难看出，在 2 个数据集上，SDT 各个特征的识别率都优于 DT。

表 1 采用 SDT、DT 方法及两者特征级联 SDT+DT 方法在 2 个数据集上的总体识别结果

方法	识别率/%	
	Hollywood2	YouTube
SDT	62.5	90.2
DT	59.9	85.7
SDT+DT	63.8	90.9



(a) Hollywood2



(b) YouTube

图 5 本文方法和传统稠密轨迹跟踪方法对 2 个数据集上的 4 个特征的识别率比较

为了进一步验证本文行为识别方法的有效性，将其和现有文献中的稠密轨迹跟踪法[1]及其他改进方法[2,3,5,17-19]进行比较。表 2 列出了本文与文献[1-3,5,17]在 Hollywood2 数据集上的最优识别结果，通过比较可以看出，本文方法的识别率虽然稍低于文献[2]中的方法，但明显高于其他文献中的方法。本文方法与文献[1,17-19]方法在 YouTube 数据集上的最优识别结果如表 3 所示，显然本文方法获得了最高的识别率。

表 2　本文方法与 5 种现有文献方法在 Hollywood2 数据集上的识别率比较

方法来源	文献[1]	文献[2]	文献[3]	文献[5]	文献[17]	本文
识别率/%	59.9	64.3	62.5	61.9	60.5	63.8

表 3　本文方法与 4 种现有文献方法在 YouTube 数据集上的识别率比较

方法来源	文献[1]	文献[17]	文献[18]	文献[19]	本文
识别率/%	85.4	86.1	87.6	90.8	90.9

4　结　论

本文针对稠密轨迹行为识别法存在的问题，采用一种两阶段显著性检测方法获取视频中的行为区域，并提取行为区域轨迹用于行为识别。第 1 阶段通过低秩矩阵恢复算法检测初始显著性，并据此将子视频划分为候选前景和绝对背景；第 2 阶段利用稀疏表示算法获取候选前景的细化显著性。这种检测方法能够以更高的对比度突出行为区域，抑制背景区域。此外，以子视频和块为基础，考虑了显著性时空相关性，增强了检测到的行为区域的时空连续性，有利于提高轨迹的连续性和完整性。实验结果表明，无论视频是否包含相机运动，本文方法都能较好地检测其中的行为区域，获取的行为识别结果优于传统稠密轨迹法和大部分改进方法。

参考文献：

[1] WANG H, KLASER A, SCHMID C, et al. Dense trajectories and motion boundary descriptors for action recognition [J]. International Journal of Computer Vision, 2013, 103: 60-79.

[2] WANG H, SCHMID C. Action recognition with improved trajectories [C]// Proceedings of IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2013: 3551-3558.

[3] JAIN M, JEGOU H, BOUTHEMY P. Better exploiting motion for better action recognition [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2013: 2555-2562.

[4] WANG X, QI C. Saliency-based dense trajectories for action recognition using low-rank matrix decomposition [J]. Journal of Visual Communication & Image Representation, 2016, 47: 361-374.

[5] VIG E, DORR M, COX D. Space-variant descriptor sampling for action recognition based on saliency and eye movements [C] // Proceedings of 12th European Conference on Computer Vision. Berlin, Germany: Springer, 2012: 84-97.

[6] SOMASUNDARAM G, CHERIAN A, MORELLAS V, et al. Action recognition using global spatio-temporal features derived from sparse representations [J]. Computer Vision and Image Understanding, 2014, 123(7): 1-13.

[7] 方志明，崔荣一，金璟璇. 基于生物视觉特征和视觉心理学的视频显著性检测算法 [J]. 物理学报，2017，66(10)：319-332.
FANG Zhiming, CUI Rongyi, JIN Jingxuan. Video saliency detection algorithm based on biological visual feature and visual psychology theory [J]. Acta Physica Sinica, 2017, 66(10): 319-332.

[8] LIU Z, LI J, YE L, et al. Saliency detection for unconstrained videos using superpixel-level graph and spatiotemporal propagation [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 27(12): 2527-2542.

[9] 陈昶安，吴晓峰，王斌，等. 复杂扰动背景下时空特征动态融合的视频显著性检测 [J]. 计算机辅助设计与图形学学报，2016，28(5)：802-812.
CHEN C A, WU X F, WANG B, et al. Video saliency detection using dynamic fusion of spatial-temporal features in complex background with disturbance [J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(5): 802-812.

[10] CANDES E J, LI X, MA Y, et al. Robust principal component analysis? [J]. Journal of the ACM, 2011, 58(3): 11.

[11] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

（下转第 44 页）