Journal of Computer Applications
计算机应用,2012,32(6):1578-1580,1608
ISSN 1001-9081
CODEN JYIIDU
2012-06-01
http://www.joca.cn

文章编号:1001-9081(2012)06-1578-03
doi:10.3724/SP.J.1087.2012.01578

基于级联结构的人体动作识别方法

彭江平*
(湖南大学 工商管理学院,长沙 410006)
(*通信作者电子邮箱 382863070@qq.com)

摘　要:提出一种基于级联结构的人体动作识别方法:针对 Dollar 时空兴趣点检测器易受图像噪声、摄像机运动与缩放等因素影响产生伪兴趣点的问题,提出了一种基于轨迹差异度的兴趣点筛选方法,有效避免了引入背景中的伪兴趣点,提高了人体运动特征提取的准确度;采用规范切与最小冗余最大相关(mRMR)准则对词袋模型生成的特征向量进行自动特征选择,同时建立一个用于分类的级联结构,在识别各类不同动作时选择不同的特征子集,使得分类器使用的特征更具区分性。在 KTH 人体运动测试集上实验,验证了该方法能提高动作识别的准确度。

关键词:人体动作识别;时空兴趣点筛选;规范切;最小冗余最大相关准则;级联分类器
中图分类号:TP391.41　　**文献标志码**:A

Human action recognition based on cascaded structure

PENG Jiang-ping*
(*School of Business Administration, Hunan University, Changsha Hunan* 410006, *China*)

Abstract: A human action recognition method based on cascaded structure was proposed in this paper. Firstly, a trajectory-based method was proposed to select the interest points detected by the Dollar detector, which was sensitive to image noise, camera movement and zooming. Therefore, the pseudo interest points in the background could be effectively excluded and the extracted features would be more relevant to action recognition. Secondly, an automatic feature selection method based on the combination of normalized cuts and mRMR criteria was used to determine a subset of the words generated by the Bag-of-Words model and construct a cascaded structure for action recognition. The purpose was to make the feature used by the cascaded structure more distinct. Lastly, the experimental results validate the contribution to the improvement of accuracy in human action recognition.

Key words: human action recognition; space-time interest point selection; normalized cut; minimal-Redundancy-Maximal-Relevance (mRMR) criteria; cascaded classifier

0　引言

人体动作识别是计算机根据来自摄像机的视频数据,通过视觉信息的处理和分析,识别人体的动作,并结合相应的环境信息,对人体动作的目的及所传递的信息进行语义描述。利用计算机进行人体动作识别不仅可以减少传统人工识别的工作量,而且随着计算机识别效率和精确度的提升所带来的经济效益和社会效益也将日益增加。目前,人体动作的识别技术在很多领域都有着广泛的应用,例如视频监控、人机交互、基于内容的图像存储和检索、视频会议、动作性能分析、虚拟现实等领域。

人体动作识别技术的研究在国内外已经开展多年,目前主要的研究方法大致可以分为以下几种:基于目标跟踪的方法[1-2]、基于光流的方法[3-4]、基于形状模板匹配的方法[5-7]和基于时空兴趣点分析的方法[8-9]。基于目标跟踪和形状模板匹配的方法在提取人体动作的全局特征之前首先要得到人体的轮廓或者人体动作的形状模板,而轮廓提取过程和人体动作形状模板构造过程容易受到噪声和背景变化的干扰,系统鲁棒性相对较差。基于光流的方法利用像素间的光流信息进行人体动作识别,但这类方法容易受到噪声以及光照强度变化的干扰。基于时空兴趣点分析的方法通过滤波、系数变换等数学方法,求出视频滤波响应或者系数变换的极值点,并将极值点作为人体动作的时空兴趣点,从中提取人体动作时空特征作为识别的依据。

本文针对 Dollar 时空兴趣点检测器易受图像噪声、摄像机运动与缩放等因素影响而产生伪兴趣点的问题,提出了一种基于轨迹差异度的兴趣点筛选方法,有效避免了引入背景中的伪兴趣点,提高了人体运动特征提取的准确度。其次,提出基于规范切与最小冗余最大相关(minimal-Redundancy-Maximal-Relevance,mRMR)准则对词袋模型生成的特征向量进行自动特征选择,在识别各类不同动作时选择不同的特征子集,使得训练分类器的特征更具区分性。在 KTH 人体运动测试集上实验,验证了改进的方法能提高动作识别的准确度。

1　人体时空兴趣点的提取

视频中的时空兴趣点定义为视频数据在空间和时间两个维度上都发生剧烈变化的地方,因此可被用来检测视频中发生的时空事件。但时空兴趣点检测器易受到背景噪声、摄像机镜头移动和伸缩的影响,在图像背景中产生伪兴趣点,影响运动人体特征的提取。因此,本文使用 Dollar 检测器[10]得到视频中的所有候选兴趣点后,利用基于轨迹差异度的方法除去候选兴趣点中的伪兴趣点,使保留点大都来自于运动人体。

收稿日期:2011-11-18;**修回日期**:2012-01-09。
基金项目:国家自然科学基金资助项目(71171076);中央高校基本科研业务青年扶持项目(11HDSK203)。
作者简介:彭江平(1967-),男,湖南湘潭人,副教授,博士,主要研究方向:计算机系统。

对于一段包含运动人体的视频片段,本文首先采用 Dollar 检测器对每一帧图像进行时空兴趣点的检测,并采用尺度不变特征变换(Scale Invariant Feature Transform,SIFT)匹配的方法[11]获得兴趣点的轨迹。对于得到的轨迹,一部分来自于背景中的兴趣点,另一部分由视频中的人体运动产生。对于视频中的一帧 f,假设有 N 条轨迹穿过该帧:$T = \{t_i\}, i = 1,2,\cdots,N$。对每一条轨迹 i,定义以帧 f 为中心,长度为 4 的轨迹段 $t_i = \{(x^i_{f-1}, y^i_{f-1}), (x^i_f, y^i_f), (x^i_{f+1}, y^i_{f+1}), (x^i_{f+2}, y^i_{f+2})\}$,则描述该轨迹段运动方向的矢量可表示为 $d_i = \{\boldsymbol{d}^i_1, \boldsymbol{d}^i_2, \boldsymbol{d}^i_3\}$,其中 $\boldsymbol{d}^i_k = (x^i_{f+(k-1)} - x^i_{f+(k-2)}, y^i_{f+(k-1)} - y^i_{f+(k-2)})$。为了度量两条轨迹段的运动差异性,可定义:

$$D_{i,j} = \sum_{k=1}^{3} \| \boldsymbol{d}^i_k - \boldsymbol{d}^j_k \|; \ i = 1,2,\cdots,N, \ j = 1,2,\cdots,N \tag{1}$$

通过计算任意两条轨迹段的运动差异性 $D_{i,j}$,可得到一个运动差异性矩阵 $\boldsymbol{M}_1 = \{D_{i,j}\}_{i=1,\cdots,N, j=1,\cdots,N}$。然后累加矩阵中的每一行元素,可得 $m_i = \sum_{j=1}^{N} D_{i,j}$。这里 m_i 衡量了以第 f 帧为中心帧,长度为 4 帧的第 i 条轨迹段与其余轨迹段的差异度。通过实验我们发现,当某条轨迹段的差异度较小时,该轨迹一般由背景兴趣点产生,这是由于背景中伪兴趣点运动轨迹一般趋于一致,因此将其去除。按式(2)可计算一个自适应的阈值:

$$Threshold_f = \frac{\kappa}{N} \sum_{i=1}^{N} m_i \tag{2}$$

其中 κ 在实验中设定为 1.2。当第 $i(i = 1,2,\cdots,N)$ 条轨迹的差异度 m_i 小于阈值 $Threshold_f$ 时,认为它是由背景兴趣点产生的,则该条轨迹段上的兴趣点将被除去,并将保留的兴趣点作为运动人体的特征点。

2　基于级联结构的动作识别

兴趣点检测器提取到的只是一系列孤立的点,而这些点的时空邻域不仅包含了与人体动作相关的信息,同时还可以增强局部特征对旋转、仿射和光照等变化的鲁棒性[10]。因此,我们按照文献[10]的方法,在检测并筛选完时空兴趣点后,提取包含该兴趣点邻域的时空立方体(cuboids),然后通过 3D SIFT 局部描述子[11]将每个时空立方体表示成一个固定维数的特征向量,采用 K 均值聚类产生码书(codebook),最后由词袋(Bag-of-Words,BoW)可将每段人体运动视频表示成一个关于码书词典的分布直方图,见图 1。

从图 1 可知,步行和慢跑有较为相似的直方图分布,拳击与挥手在分布上较为相似。直观上看,(a)、(b)两组动作之间第 1 ~ 20 列的直方图分布差异较大,因此选择第 1 ~ 20 列特征来分开(a)、(b)两组动作将具有较好的区分性。单独考察图(a)可知,这两种动作第 1 ~ 20 列的特征分布较为相似,而第 20 ~ 30 列特征分布的差异较大,因此选用第 20 ~ 30 列特征区分步行与慢跑这两种动作将比选用第 1 ~ 20 列特征更为合适。由此可见,为了区分不同的动作应当选择不同的特征子集,使得训练动作分类器的特征更具代表性。

为了对每类人体运动视频识别时选用最具区分性的直方图子集,这里构造一个二叉树型级联分类结构,如图 2 所示。在级联结构的每一级,采用谱聚类的方法将待分类的一组动作集分成两个子集 S 和 $\bar{S}$。在子集 S 中,两两动作的相似度大于与子集 $\bar{S}$ 中任何一类动作的相似度。在级联结构叶节点所示的集合中只包含一类动作。当分类进行到叶节点时,叶节点对应的类型即为待识别动作的类别。

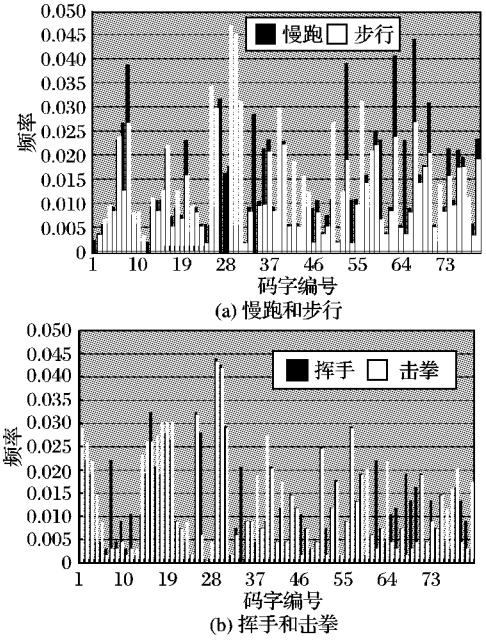


图 1　4 种动作的码书词典直方图

具体说来,在级联结构的第 1 级,对于待分类的 N 类动作 $S_1 = [a_1, a_2, \cdots, a_N]$,首先计算 N 类动作两两之间的相似度,并将其写成 $N \times N$ 的相似度矩阵:

$$\boldsymbol{M}_2 = \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{pmatrix}$$

这里 S_{ij} 表示第 i 类动作与第 j 类动作的相似度,可表示为:

$$S_{ij} = 1 - \| \boldsymbol{p}_i - \boldsymbol{p}_j \| \tag{3}$$

其中:$\boldsymbol{p}_i$ 代表第 i 类动作的码书均值直方图,$\| \boldsymbol{p}_i - \boldsymbol{p}_j \|$ 代表第 i 类和第 j 类码书直方图的欧氏距离。获得动作的相似度矩阵以后,采用基于谱图理论的聚类方法规范切[12]可将 N 类动作集合 S 按照上述要求分成两个子集。该过程一直循环,直到每个子集只含有一类动作。

图 2 所示级联结构的每个分叉处对应一个二分类器,训练每个二分类器所用的特征(对应于码书直方图中的列)通过 mRMR 准则[13]来筛选。mRMR 准则是一种基于互信息的特征选择算法,通过它可以选出特征冗余度最低同时分类错误率也最低的一个特征子集。

3　实验结果和分析

为了验证本文方法的有效性,我们在目前被普遍使用的 KTH 人体动作库上进行实验。KTH 集包含了 6 种人体动作:慢跑、跑、行走、击拳、挥手和击掌。它由 25 个人将这 6 种动作在 4 种不同的场景中重复数次,一共包括 2 391 段视频序列。每个视频的帧率都为 25 fps,分辨率为 160 × 120,持续时间为4 s。

图 3 显示了采用 Dollar 检测器进行时空兴趣点检测后,未经过伪兴趣点去除和经过去除后的实验结果对比。为了表示方便,将对连续多帧图像进行兴趣点检测的结果累计在了同一帧上。由图 3(a)可知,由于摄像机晃动、缩放等原因,检测到的一部分时空兴趣点为背景点,这样会对后续动作识别产生不利影响。图 3(b)表明本文基于轨迹差异度的伪兴趣点去除可有效排除落在背景中的伪兴趣点,使提取到的时空兴趣点大都来自运动人体。

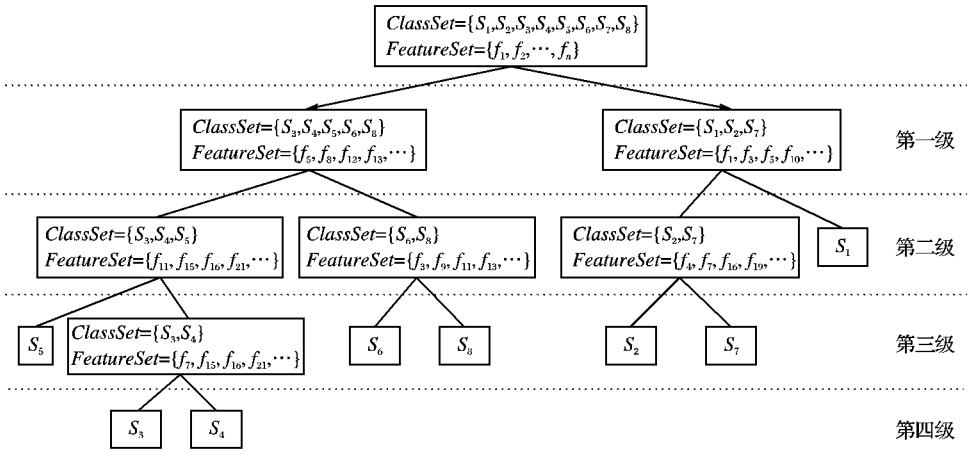
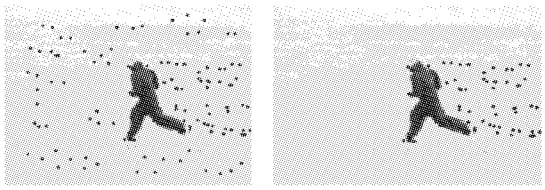


图 2 级联结构示例



(a) 未筛选的实验结果 (b) 筛选后的实验结果

图 3 兴趣点筛选对比实验

接下来本文对采用级联结构进行特征选择能否提高动作识别准确度做了对比实验。由于本文只是为了验证:当采用同一种二分类器时,级联结构比非级联结构的动作识别准确度要高,因此级联结构中的二分类器类型任意,但不同二分类器下的识别准确度会有所不同。本文实验中选用支持向量机(Support Vector Machine,SVM)作为级联结构中的二分类器。

利用本文第 2 章介绍的方法对 KTH 动作库建立的级联结构如图 4 所示,该级联结构由三级构成,每级通过不同的分类器来分开不同的动作。例如,第一级左边的集合包含的均是上半身的手臂运动,右边的集合包含的则是下半身的腿部运动。分类器分开动作的相似度逐层递增。采用级联结构与直接采用多类 SVM 的平均识别准确度如图 5 所示。为了不失一般性,我们统计了识别率在不同码书大小下的实验结果。从图 5 可知,级联结构下的平均动作识别率普遍高于非级联结构的平均识别率。级联结构下对 KTH 集中 6 种动作识别率的混淆表见表 1。从表中可知,每种动作平均识别率均在 85% 以上,平均识别率为 89.8%,达到了较理想的识别效果。

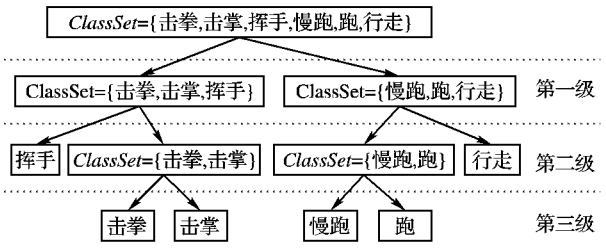


图 4 KTH 集下的级联结构

表 1 级联结构下对 KTH 集上每种动作的识别率 %

动作	击拳	击掌	挥手	慢跑	跑	行走
击拳	86	7	0	0	0	7
击掌	5	92	3	0	0	0
挥手	2	9	89	0	0	0
慢跑	0	0	0	86	0	14
跑	0	0	0	4	93	3
行走	0	0	0	7	0	93

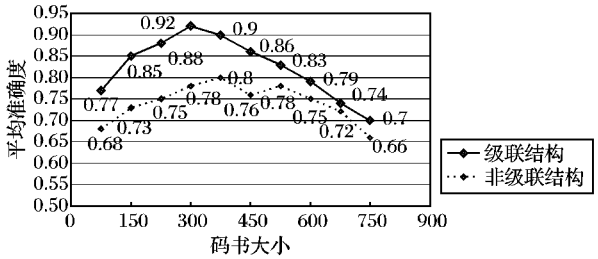


图 5 级联结构与非级联结构的实验结果对比

4 结语

基于时空兴趣点的人体动作识别是当前的研究热点。由于摄像机晃动和缩放、图像噪声等因素影响,时空兴趣点检测器容易在背景图像中产生伪兴趣点,从而影响运动人体特征的提取,降低了动作识别的准确度。本文通过一种基于轨迹差异度的方法对候选兴趣点集进行筛选,有些避免了引入背景中的伪兴趣点,提高了运动人体特征提取的准确度。同时利用规范切与 mRMR 准则自动建立级联的分类结构,在识别各类不同动作时选择不同的特征子集。实验结果表明,本文方法能提高动作识别的准确度。

参考文献:

[1] RAMANAN D, FORSYTH D A. Automatic annotation of everyday movements[C]// 17th Annual Conference on Neural Information Processing Systems. Vancouver: NIPSF, 2003: 77-84.

[2] SHEIKH Y, SHEIKH M, SHAH M. Exploring the space of a human action[C]// IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2005: 144-149.

[3] EFROS A, BERG A, MORI G, *et al*. Recognizing action at a distance[C]// IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2003: 726-733.

[4] FATHI A, MORI G. Action recognition by learning midlevel motion features[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2008: 726-733.

[5] GORELICK L, BLANK M, SHECHTMAN E, *et al*. Actions as space-time shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2247-2253.

[6] KE Y, SUKTHANKAR R, HEBERT M. Efficient visual event detection using volumetric features[C]// The 10th IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2005: 166-173.

(下转第 1608 页)

2.3 能量消耗

在资源受限的无线传感器网络中,存储需求显得尤为重要。为了降低存储消耗,本文方案采用文献[12]中的编码方案,将矩阵 L 的行向量和矩阵 U 的列向量分为非零部分和零部分,在计算每个节点存储消耗时,计算零元素的个数加上非零元素部分占有的存储空间,由于 LU 矩阵是三角矩阵,所以整个网络的存储消耗可大致节约50%。

将本文方案与文献[8]比较,由定义1可知,本文方案利用 $[m(m+1)/2] \times |S|$ 个密钥生成所需的密钥矩阵空间,子矩阵对的阶数为 m。文献[8]采用同样多的密钥形成 LU 矩阵,若矩阵的阶数为 n,则满足 $[m(m+1)/2] \times |S| = n(n+1)/2$,则在建立直接密钥的过程中,本文方案只需传输长度为 m 的列向量即可,比文献[8]传输消耗少得多。具体比较见表2。

表2 本文方法与文献[8]方法传输消耗比较

m	$\|S\|$	传输列向量长度		传输消耗节省率/%
		本文方案	文献[8]方法(近似值)	
4	50	4	31	87.0
4	100	4	44	91.0
6	50	6	45	86.7
6	100	6	64	90.6

由表2可知,本文节点的传输消耗比仅采用 LU 矩阵法的文献[8]的节点的传输消耗最多可节省90%以上。主要是由于本文将分组部署策略引入到 LU 矩阵空间中,通过子矩阵对的分组部署,大大减少了节点传输向量长度,从而节约了传输消耗,这对资源有限的无线传感器网络应用是行之有效的。分析表2可进一步得知:在条件相同的情况下,矩阵空间的矩阵对个数越多,传输消耗节省的越多;子矩阵对元素的阶数越小,传输消耗节省的越多。

3 结语

本文着重研究基于部署策略的 LU 矩阵空间密钥管理方案,该方案在通过对部署区域的合理分区及子密钥矩阵空间元素的合理选择,较好地解决了 LU 矩阵密钥管理的信息泄漏问题,提高了无线传感器网络节点的抗捕获攻击能力,并有效提高节点的通信效率,节点通信开销比文献[8]最多可节约90%以上,较好地达到了无线传感器网络的连通性、安全性和能量消耗的平衡点。但是,本文研究仍需进一步拓展和深入,并争取能在实际的无线传感器网络模型中实现应用。

参考文献:

[1] REN HENG, SUN XINGMING, RUAN ZHIQIANG, *et al*. An efficient scheme against node capture attacks using secure pairwise key for sensor networks[J]. Information Technology Journal, 2011, 10(1):71-79.

[2] 袁斑,马建庆,钟亦平,等.基于时间部署的无线传感器网络密钥管理方案[J].软件学报,2010,21(3):516-527.

[3] 马春光,张秉政,孙原,等.基于按对平衡设计的异构无线传感器网络密钥预分配方案[J].通信学报,2010,31(1):37-43.

[4] CHOW CHI-YIN, MOKBEL M F, HE TIAN. A privacy-preserving location monitoring system for wireless sensor networks[J]. IEEE Transactions on Mobile Computing, 2011, 10(1):94-107.

[5] ESCHENAUER L, GLIGOR V D. A key-management scheme for distributed sensor networks[C]// Proceedings of the 9th Association for Computing Machinery Conference on Computer and Communications Security. New York: ACM, 2002:41-47.

[6] LIU DONGGANG, NING PENG. Location-based pairwise key establishments for static sensor network[C]// Proceedings of the 1st Association for Computing Machinery Workshop on Security of Ad-Hoc and Sensor Networks. New York: ACM, 2003:72-82.

[7] DU WENLIANG, DENG JING, HAN Y S, *et al*. A key management scheme for wireless sensor networks using deployment knowledge[C]// Proceedings of the IEEE Computer and Communication Societies. Piscataway: IEEE Press, 2004:586-597.

[8] CHOI S J, YOUN H Y. An efficient key predistribution scheme for secure distributed sensor networks[C]// 2005 International Federation Information Processing International Conference on Embedded and Ubiquitous Computing. Berlin: Springer, 2005: 1088-1097.

[9] ZHU BO, ZHENG YANFEI, ZHOU YAOWEI, *et al*. Cryptanalysis of LU decomposition-based key pre-distribution scheme for wireless sensor networks[EB/OL].[2011-10-20]. http://eprint.iacr.org/2008/411.pdf.

[10] DAI HANG-YANG, XU HONG-BING. Key predistribution approach in wireless sensor networks using LU matrix[J]. IEEE Sensors Journal, 2010, 10(8):1399-1409.

[11] 余旺科,马文平,王淑华.基于部署信息的无线传感器网络密钥预分配[J].华中科技大学学报:自然科学版,2010,38(11):51-54.

[12] TRAN T D, AL-SAKIB K P, CHOONG S H. A resource-optimal key pre-distribution scheme with enhanced security for wireless sensor networks[C]// Proceedings of the 9th Asia-Pacific International Conference on Network Operations and Management: Management of Convergence Networks and Services. Berlin: Springer-Verlag, 2006:546-549.

(上接第1580页)

[7] YILMAZ A, SHAH M. Actions sketch: a novel action representation[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2005: 984-989.

[8] GILBERT A, ILLINGWORTH J, BOWDEN R. Scale invariant action recognition using compound features mined from dense spatio-temporal corners[C]// The 10th European Conference on Computer Vision. Berlin: Springer-Verlag, 2008: 222-233.

[9] SAVARESE S, POZO A D, NIEBLES J, *et al*. Spatial temporal correlations for unsupervised action classification[C]// IEEE Workshop on Motion and Video Computing. Piscataway: IEEE Press, 2008: 1395-1402.

[10] DOLLAR P, RABAUD V, COTTRELL G, *et al*. Behavior recognition via sparse spatio-temporal features[C]// The 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Washington, DC: IEEE Computer Society, 2005: 65-72.

[11] SUN JU, WU XIAO, YAN SHUICHENG, *et al*. Hierarchical spatio-temporal context modeling for action recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009:2004-2011.

[12] SHI J, MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8):888-905.

[13] PENG HANCHUAN, LONG FUHUI, DING C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern analysis and Machine Intelligence, 2005, 27(8):1226-1238.