

多源信息融合的双人交互行为识别算法研究

金壮壮¹, 曹江涛¹, 姬晓飞²

(1. 辽宁石油化工大学 信息与控制工程学院, 辽宁 抚顺 113001;
2. 沈阳航空航天大学 自动化学院, 辽宁 沈阳 110136)

摘要: 人体及人体的运动均是三维信息, 而传统的基于 RGB 视频的双人交互行为的特征描述方法由于缺少深度信息导致其特征描述的区分度较低。根据 RGB 视频和深度视频各自优点和具有的互补特性, 提出一种多源信息融合的双人交互行为识别算法。该算法首先采用时空兴趣点和词袋模型结合的方法对 RGB 视频进行特征表示。然后采用方向梯度直方图对深度视频帧进行特征表示, 并引入关键帧统计特征对深度视频进行直方图特征表示。最后, 使用最近邻分类器分别对两种视频特征进行分类识别, 通过加权融合两类视频的识别概率实现交互行为的识别。实验结果表明, 引入深度信息可以大大提高双人交互行为识别的准确率。

关键词: 动作识别; 时空兴趣点; 方向梯度直方图; 加权融合; 双人交互; 词袋模型

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2018)10-0032-05

doi: 10.3969/j.issn.1673-629X.2018.10.007

Research on Human Interaction Recognition Algorithm Based on Multi-source Information Fusion

JIN Zhuang-zhuang¹, CAO Jiang-tao¹, JI Xiao-fei²

(1. School of Information and Control Engineering, Liaoning Shihua University, Fushun 113001, China;
2. School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: The human and the human actions are three-dimensional information. However, the traditional feature description method of human interaction based on RGB video without the depth information lead to a lower discriminative ability. Considering the respective advantages of the RGB image and the depth image and the characteristics of information complementarity, we propose a novel multi-source information fusion algorithm of human interaction recognition. Firstly, the spatial-temporal interest point and the BOW model are used to represent RGB video. Secondly, the histogram of oriented gradient is used to describe the frames of depth video. In addition, the histogram representation is obtained by using key frame statistical feature. Finally, the nearest neighbor classifier is used to classify two kinds of video features respectively, then human interaction recognition is achieved by weighted fusing recognition probability of two types of video. The experiment shows that the accuracy of the human interaction recognition is improved by combing the depth information.

Key words: action recognition; spatio-temporal interest point; histogram of oriented gradient (HOG); weighted fusion; human interaction; bag of word (BOW)

1 概述

在信息技术高速发展的今天, 机器视觉越来越受到研究者的关注, 其中双人交互行为识别算法研究已成为热点问题。相比单人行为模型, 双人交互行为更为复杂, 因此, 如何有效地提取运动特征和合理地建立交互运动模型是双人交互行为识别与理解的重要研究

内容。国内外科研工作者们已经展开了相关项目的研究^[1-5]。对双人交互行为进行自动的识别不仅对计算机视觉、人工智能和模式识别的理论研究具有很高的价值, 而且在解决公共场合暴力冲突和实现智能视频监控等方面具有现实意义。

在过去的十几年中, 双人交互行为识别主要针对

收稿日期: 2017-10-19

修回日期: 2018-03-05

网络出版时间: 2018-05-16

基金项目: 辽宁省自然科学基金(201602557); 辽宁省高等学校优秀人才支持计划项目(LJQ2014018, LR2015034)

作者简介: 金壮壮(1991-), 男, 硕士, 研究方向为模式识别、视频分析; 曹江涛, 教授, 硕导, 研究方向为智能方法及其在工业控制和视频信息处理上的应用; 姬晓飞, 副教授, 硕导, 研究方向为视频分析与处理、模式识别。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20180515.1702.070.html>

RGB 视频序列进行。Kong 等^[6]提出一种基于补丁感知模型的双人交互行为识别方法,该方法从包含双人交互的视频中提取时空兴趣点,构造视觉词汇直方图,然后从直方图中采样非重叠的 3D 补丁,使用提出的补丁感知模型来推断潜在补丁变量的标签并对交互动作进行分类识别。Zhang 等^[7]使用鸟瞰图摄像机的空间线索来识别发生交互的时间间隔,然后从分段视频中检测时空兴趣点,利用 K-means 构建视觉码本。在视觉码本上投影每个视频,构造视觉词汇直方图来实现交互识别。这些算法具有较好的性能,但 RGB 视频将人体及人体运动投影到二维平面,导致深度信息的缺失,因此对于复杂交互动作识别的准确性很难得到保证。

近年来,微软推出一种可以同时拍摄视觉图像(RGB 图像)与深度图像的 Kinect 设备,大大降低了深度信息的获取成本,在智能监控、人机交互等领域受到了重视,一些研究者将其引入到双人交互行为识别的研究中。Yun 等^[8]采用文献[9]中的几何特征提取方法并设计多种相关的距离特征(如关节距离、关节运动、平面特征和速度特征)来进行双人交互识别,发现关节距离特征和关节运动特征在双人交互行为识别上优于平面特征和速度特征。该类方法依赖于 Kinect

设备对关节点信息的准确估计,对于复杂行为识别的准确率不高。Ji 等^[10]通过提取关节点运动与关节点距离特征来进行交互身体部位描述,然后构建对比特征分布模型(contrastive feature distribution model, CFDM)实现双人交互行为识别。该方法依赖骨架模型的构建及关节点信息之间的准确联系,尽管识别准确率较高,但其计算复杂度也较高,不适于实时应用。

目前基于 RGB 视频和深度视频的双人交互行为识别研究均取得了一定的进展,但结合两者进行的研究还处于起步阶段。Ni 等^[11]对传统时空兴趣点检测算法做了相应改进,提出了一种基于深度信息分层的时空兴趣点特征描述算法。其算法主要思想是把深度图像按灰度级分成 M 层通道,将 RGB 视频构建的 K 维视觉词典与每一层通道进行匹配,生成 $M \times K$ 维特征矢量。该算法在一定程度上提高了识别的准确性,但当 K 值较大时才能得到较好的效果,且特征矢量维数增加会降低动作识别效率。

根据以上分析,综合考虑到基于 RGB 图像与深度图像各自的优点且具有信息互补的特性,文中提出一种多源信息融合的双人交互行为识别算法。算法框图如图 1 所示。

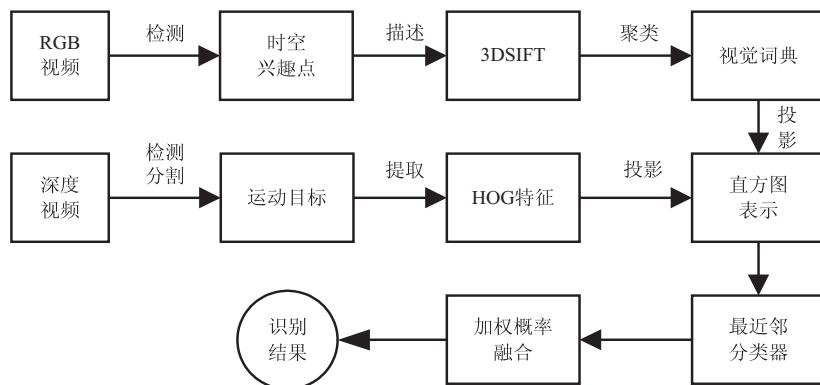


图 1 算法结构框图

算法实现步骤如下:

(1) RGB 视频的特征表示:时空兴趣点特征(spatial-temporal interest point, STIP)在 RGB 视频上已成功应用,因此在 RGB 视频方面首先进行时空兴趣点检测,用三维尺度不变特征转换(3-dimensional scale invariant feature transform, 3DSIFT)对兴趣点进行描述,采用 BOW 模型对 RGB 视频进行直方图表示。

(2) 深度视频的特征表示:在深度视频方面,首先进行深度图像检测分割处理。考虑到一般情况下前景和背景之间在深度方向存在一定的距离,因此在深度图像中,前景的边缘信息明显,选取方向梯度直方图(histogram of oriented gradient, HOG)特征对每个视频帧中的检测区域进行特征表示,然后采用关键帧统计

特征对整个深度视频进行表示。

(3) 决策级融合识别:使用最近邻分类器计算待测试视频与动作模板的相似性概率,之后对 RGB 图像上提取时空兴趣点特征的识别结果与深度图像上提取 HOG 特征的识别结果进行加权融合,得到待测视频的最终识别结果。

2 RGB 视频特征表示

在人类交互行为视频里,时空兴趣点能够用较少的信息量正确地定位视频序列中具有明显运动的区域,对环境的变化、局部的遮挡具有较强的鲁棒性^[12-13]。基于时空兴趣点表征行为的算法被广泛应用于人类行为识别领域,因此文中使用时空兴趣点与

视觉词袋 (bag of word, BOW) 模型结合的方法对 RGB 视频进行特征表示。

如图 2 所示,使用时空兴趣点和词袋模型相结合的双人交互行为表征算法由时空兴趣点检测、特征描述、词典建立三个部分组成。

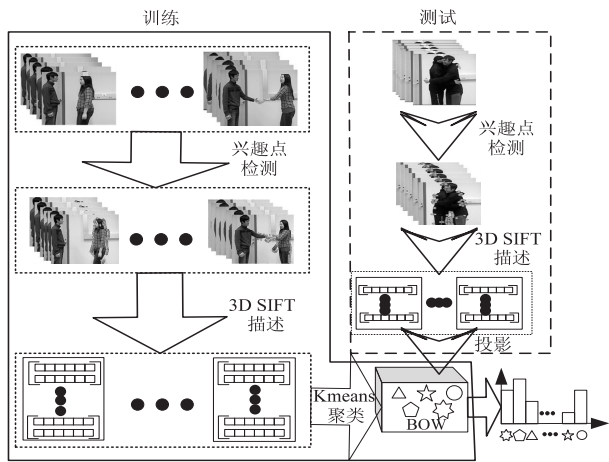


图 2 生成 BOW 描述符的图形表示

2.1 时空兴趣点检测

广泛使用的兴趣点检测方法是由 Dollars 等^[14]提出的基于 Gabor 滤波器和高斯滤波器组合计算函数响应值的方法。将空间和时间两个单独的线性滤波器应用在检测器上,从视频序列中提取出丰富的时空兴趣点,用来充分捕捉视频序列中的人体行为特征。这里把它应用在彩色视频双人交互局部特征提取上,其表达式为:

$$R = (I * g(\sigma) * h_{ev})^2 + (I * g(\sigma) * h_{od})^2 \quad (1)$$

其中, $g(x, y; \sigma)$ 是二维高斯平滑核函数,用于空间域滤波。

$$g(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp(-(x^2 + y^2)/2\sigma^2) \quad (2)$$

其中, h_{ev} 和 h_{od} 是一维 Gabor 函数的正交分量,用于时间域滤波。

$$h_{ev}(t; \tau, \omega) = -\cos(2\pi\omega t) e^{-t^2/\tau^2} \quad (3)$$

$$h_{od}(t; \tau, \omega) = -\sin(2\pi\omega t) e^{-t^2/\tau^2} \quad (4)$$

其中, σ 和 τ 分别对应空间和时间尺度。

被检测为兴趣点需满足两个条件,响应函数 R 大于设定阈值且在某一邻域内取得局部极大值,阈值大小的选取可以控制检测出的兴趣点数目^[15]。

2.2 3D SIFT 描述时空兴趣点

采用 3D SIFT 方法进行兴趣点描述,步骤如下:

(1) 在兴趣点周围的邻域内提取时空立方体并将其划分为固定大小的单位子立方体;

(2) 使用多面球计算每个单位立方体的时空梯度直方图;

(3) 组合所有单位立方体直方图,形成时空兴趣

点的 3D SIFT 描述符^[16]。

文中 $X \times Y \times Z$ 像素大小的立方体被划分为 M 个子立方。采用 P 个面对 Q 个梯度方向进行描述,因此每个点的特征维数是 $R \times S$,用以描述交互行为的时空兴趣点特征^[17]。

2.3 词典构建

使用简单有效的 K 均值聚类算法对所有特征向量进行聚类,其处理过程是:先在数据集中随机选择 K 个样本作为聚类中心,根据一定的相似性度量把所有样本划分到与之距离最近的聚类中心所代表的类中,形成 K 个聚类,然后对这 K 个聚类重新计算聚类中心,并按照新的聚类中心重新划分样本类别,如此迭代进行下去,直到式 5 准则函数收敛为止。

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (5)$$

其中, E 为所有研究对象的平方误差总和; p 为空间的点,即数据对象; m_i 为簇 C_i 的平均值。

把每个聚类中心作为词典中的一个单词,每个特征向量用与之距离最近的单词表示,然后对词典中的单词在视频中出现的频率进行统计,构建视频的直方图表示。

3 深度视频特征表示

深度图像也称为距离影像,是指将图像采集器到场景中各点的距离(深度)作为像素值的图像^[18]。当人体前景与背景存在一定的距离时,深度图像可以通过灰度值信息直观地体现。考虑到 HOG 可以较好地对人体周围的边缘信息进行提取和表示,因此选用 HOG 特征进行深度视频全局表示。文中采用帧差法^[19]检测深度图像中的运动目标。

HOG 特征^[20]的构建通过计算和统计图像局部区域的梯度方向直方图来实现。针对梯度进行提取操作,不但可以捕获轮廓和纹理信息,还可以减小光照变化的影响。像素点 (x, y) 横坐标和纵坐标方向的梯度可表示为:

$$\begin{cases} G_x(x, y) = H(x+1, y) - H(x-1, y) \\ G_y(x, y) = H(x, y+1) - H(x, y-1) \end{cases} \quad (6)$$

其中, $G_x(x, y)$ 、 $G_y(x, y)$ 、 $H(x, y)$ 分别表示输入图像中像素点 (x, y) 处的水平方向梯度、垂直方向梯度和像素值。

像素点 (x, y) 处的梯度幅值和梯度方向分别表示为:

$$\begin{cases} G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ \alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \end{cases} \quad (7)$$

将运动目标所在区域进行 HOG 特征提取。即梯

度图像均等划分为 $P \times Q$ 个不重叠子区域,并计算每个区域中像素点的梯度对 K 个不同方向的贡献权重大小,将其叠加到所有的梯度方向上,构建梯度方向直方图,如图 3 所示。最终得到每帧图像中运动目标的 $P \times Q \times K$ 维特征向量。

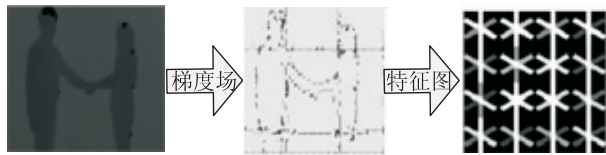


图 3 HOG 特征图形表示

为了与兴趣点特征进行有效融合,使用关键帧统计特征对深度视频进行表示。即采用 K -means 聚类方法,对训练视频的 HOG 特征生成关键帧特征库。然后根据相似度量函数对一个待测视频中的所有帧特征在关键帧特征库中出现的频率进行统计,得到深度视频的统计直方图表示^[21]。

4 分类器设计

最近邻分类器是一种结构简单、识别效果良好的识别方法^[22]。文中采用最近邻分类器对两种视频特征进行识别,并融合两种特征的识别概率对交互动作进行最终识别,具体方法如下:

获得 RGB 视频与深度视频的识别概率后,通过加权融合的方式可以得到融合后的识别概率及结果:

$$P_{\text{Final}} = w_1 \times P_{\text{rgb}} + w_2 \times P_{\text{depth}} \tag{8}$$

其中, P_{Final} 为加权融合后的最终识别概率; P_{rgb} 和 P_{depth} 分别为 RGB 视频和深度视频的识别概率; w_1 、 w_2

为 RGB 视频和深度视频识别概率的加权参数。

5 实验与结果分析

5.1 数据库

实验采用公开的 SBU Kinect interaction^[8] 双人交互动作视频数据库。该数据库是运用微软 Kinect 传感器创建的一个拥有深度图像、彩色图像和骨架图像的双人交互动作数据库。数据库记录了八类双人交互动作 (approaching, departing, kicking, punching, pushing, hugging, shaking hands, exchanging), 共有七人在同一实验室环境中参与行为动作拍摄,每一类动作都由不同的动作执行人完成,整个数据库有 260 组交互动作。库内的行为不仅是非周期性行为,而且不乏有非常相似的身体动作,因此对库中的行为进行识别具有一定的挑战性。

5.2 不同特征识别结果

在实验过程中,每类动作随机选取 10 个视频作为测试数据,余下的视频作为训练数据。参数选取如下:兴趣点附近 $12 \times 12 \times 12$ 像素大小的立方体被划分为 2 个子立方,采用 32 个面对 32 个梯度方向进行描述,即每个点的特征是 256 维用以描述文中交互行为的时空兴趣点特征。在深度图像上,将梯度图像均等划分为不重叠子区域,每一幅运动区域做 4×4 的分割,方向个数为 12。在进行决策级融合之前,使用最近邻算法计算测试视频动作模板的相似性概率,为确定决策级融合时加权数值的分配提供参考识别结果。

表 1 不同特征的识别结果

特征	动作识别率/%								总识别率/%
	靠近	远离	交换	拥抱	踢腿	打拳	推搡	握手	
RGB+BOW	80	50	60	100	80	70	40	90	71.25
Depth+HOG	100	100	90	100	70	60	100	60	85.00

由表 1 可以看出,RGB 视频上采用的 BOW 特征表示相对于深度视频的 HOG 特征表示“推搡”和“远离”交互动作识别效果较差,而“握手”识别效果较好。对于双人交互动作识别,深度图像携带更多动作信息,因此基于深度图像的识别率较 RGB 视频有了大幅提高。上述结果表明,文中将 BOW 特征和 HOG 特征识别结果进行决策级融合是有意义的。

5.3 决策级融合识别结果

针对 BOW 和 HOG 各自的特征优势和特点,分别将它们最近邻识别结果进行决策级融合,根据上面得到的结果,运用遍历的方法得到最优权值分别为 0.45 和 0.55,深度视频的融合系数较高一些。最终的识别结果采用数据化后的混淆矩阵表示,如图 4 所示。

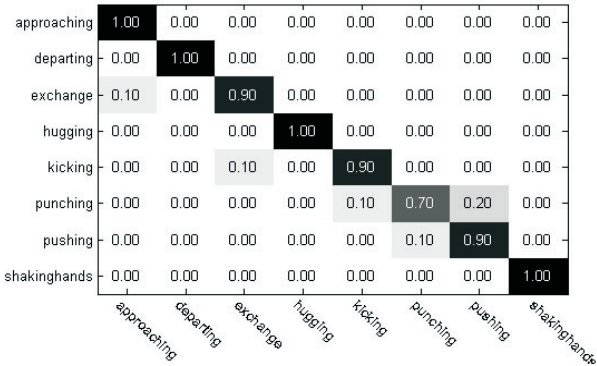


图 4 决策级融合后的混淆矩阵

从混淆矩阵可以看出,决策级融合后的正确识别率为 92.5%,远远优于单一特征的识别结果。对比单一识别结果,“靠近”、“远离”和“握手”识别效率得到

明显提高。由于数据库有些动作存在相似性,导致融合后有 20% 的“打拳”动作视频被识别成“推搡”交互动作。

5.4 实验比较

为了验证文中算法的有效性,对同样在 SBU Kinect interaction 数据库上进行实验的相关文献使用的算法与文中算法进行比较和分析,如表 2 所示。

表 2 不同方法的识别结果

算法	识别方法	识别率/%
文中	BOW+HOG	92.5
文献[8]	Raw skeleton	49.7
	Joint features	80.3
文献[23]	Rawskeleton+Contrast Mining	79.4
	Jointfeatures+Contrast Mining	86.9
文献[10]	Jointfeatures+BOW	82.5
	Jointfeatures+CFDM	89.4
文献[24]	SA-LSTM	88.0
	TA-LSTM	89.0
	STA-LSTM	91.5
文献[25]	Softmax+LCNN	92.4
	SVM+LCNN	92.8

由表 2 可以看出,文献[8,10,23-25]选取了骨架结构模型进行双人交互动作识别。从识别率上看,文中多源信息融合方法得到了较高的正确识别率,并且文献[24]中的算法需要获取人体骨架节点信息之间的关联,而且 LSTM 存在训练复杂度高、解码时延高的问题。文献[25]的 LCNN 算法存在训练需分多个阶段,步骤繁琐,速度慢,对硬件要求高的问题。而文中算法直接在 RGB 和深度图像上提取特征,简单易实现。

6 结束语

根据 RGB 图像与深度图像特性,提出一种 RGB 视频与深度视频特征在决策级加权融合的交互动作识别算法。该算法充分利用了两种视频信息各自的优点及信息互补的特性,并采取了适用于两种视频的特征表示方法。在 SBU Kinect interaction 数据库的测试结果表明,该算法对于复杂的双人交互行为取得了良好的识别结果,并且实现简单,不依赖于 Kinect 设备对骨架模型的估计,为双人交互行为识别提供了一种新的解决方案。为了进一步提高算法的准确性,下一步的研究重点是更好地构建联合特征模型,研究两种视频信息的特征级融合。

参考文献:

[1] KANTOROV V, LAPTEV I. Efficient feature extraction en-

coding and classification for action recognition [C]//Proceedings of the 2014 IEEE computer society conference on computer vision and pattern recognition. Columbus, USA: IEEE, 2014: 2593-2600.

[2] ZHANG Xinye, CUI Jinshi, TIAN Lu. Local spatio-temporal feature based voting framework for complex human activity detection and localization [C]//Proceedings of the 1st Asian conference on pattern recognition. Beijing, China: IEEE, 2011: 12-16.

[3] EL HOUDA SLIMANI K N, BENEZETH Y, SOUAMI F. Human interaction recognition based on the co-occurrence of visual words [C]//Proceedings of the 2014 IEEE conference on computer vision and pattern recognition workshops. Columbus, OH, USA: IEEE, 2014: 461-466.

[4] MUKHERJEE S, BISWAS S K, MUKHERJEE D P. Recognizing interaction between human performers using “key pose doublet” [C]//Proceedings of the 19th ACM international conference on multimedia. Scottsdale, AZ, USA: ACM, 2011: 1329-1332.

[5] 黄菲菲, 曹江涛, 姬晓飞, 等. 多特征的双人交互动作识别算法研究 [J]. 计算机科学与探索, 2017, 11(2): 294-302.

[6] KONG Yu, FU Yun. Close human interaction recognition using patch-aware models [J]. IEEE Transactions on Image Processing, 2016, 25(1): 167-178.

[7] ZHANG Bo, ROTA P, CONCI N. Recognition of two-person interaction in multi-view surveillance video via proxemics cues and spatio-temporal interest points [J]. Proceedings of SPIE, 2013, 8663: 866305.

[8] YUN K, HONORIO J, CHATTOPADHYAY D, et al. Two-person interaction detection using body-pose features and multiple instance learning [C]//Proceedings of CVPR workshops. Providence, RI, USA: IEEE, 2012: 28-35.

[9] MÜLLER M, RÖDER T, CLAUSEN M. Efficient content-based retrieval of motion capture data [J]. ACM Transactions on Graphics, 2005, 24(3): 677-685.

[10] JI Yanli, CHENG Hong, ZHENG Yali, et al. Learning contrastive feature distribution model for interaction recognition [J]. Journal of Visual Communication & Image Representation, 2015, 33: 340-349.

[11] NI Bingbing, WANG Gang, MOULIN P. RGBD-HuDaAct: a color-depth video database for human daily activity recognition [C]//IEEE international conference on computer vision workshops. Barcelona, Spain: IEEE, 2013: 1147-1153.

[12] 王佩瑶, 曹江涛, 姬晓飞. 基于改进时空兴趣点特征的双人交互行为识别 [J]. 计算机应用, 2016, 36(10): 2875-2879.

[13] LI Chuazhen, SU Bailiang, LIU Yin, et al. Human action recognition using spatio-temporal descriptor [C]//International congress on image and signal processing. Hangzhou, China: IEEE, 2014: 107-111.

[14] DOLLAR P, RABAU V, COTTRELL G, et al. Behavior

[12] BI Wei, KWOK J T. Hierarchical multilabel classification with minimum bayes risk[C]//12th international conference on data mining. Brussels, Belgium; IEEE,2012;101-110.

[13] VENS C, STRUYF J, SCHIETGAT L, et al. Decision trees for hierarchical multi - label classification [J]. Machine Learning,2008,73(2):185-214.

[14] BLOCKEEL H, BRUYNOOGHE M, DŽEROSKI S, et al. Hierarchical multi - classification [C]//Workshop notes of the KDD'02 workshop on multi-relational data mining. [s. l.]:[s. n.],2002;21-35.

[15] ALVES R T, DELGADO M R, FREITAS A A. Multi-label hierarchical classification of protein functions with artificial immune systems[C]//Brazilian symposium on bioinformatics. Berlin; Springer,2008;1-12.

[16] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining multi - label data [M]//Data mining and knowledge discovery handbook. Boston, MA; Springer,2009;667-685.

[17] 徐有正,黄 刚. 多标签图像的识别分类处理算法[J]. 计算机时代,2017(10):9-12.

[18] 郑晓雪,张大方,刁祖龙. 基于用户身份特征的多标签分类算法[J]. 计算机应用,2017,37(6):1697-1701.

[19] 姚小慧,孙国强. 基于局部和全局一致性的多标签分类算法[J]. 电子科技,2017,30(3):4-7.

[20] 王 梅. 基于多标签学习的图像语义自动标注研究[D]. 上海:复旦大学,2008.

[21] SHALEV - SHWARTZ S, SINGER Y. Efficient learning of label ranking by soft projections onto polyhedra[J]. Journal of Machine Learning Research,2006,7:1567-1599.

[22] LI Feifei, FERGUS R, PERONA P. Learning generative visual models from few training examples; an incremental Bayesian approach tested on 101 object categories[J]. Computer vision and Image understanding,2007,106(1):59-70.

[23] VAN GEMERT J C, GEUSEBROEK J M, VEENMAN C J, et al. Kernel codebooks for scene categorization[C]//European conference on computer vision. Berlin; Springer,2008;696-709.

[24] READ J. Scalable multi-label classification[D]. New Zealand; University of Waikato,2010.

[25] KOSMOPOULOS A, PARTALAS I, GAUSSIER E, et al. Evaluation measures for hierarchical classification; a unified view and novel approaches[J]. Data Mining and Knowledge Discovery,2015,29(3):820-865.

[26] GODBOLE S, SARAWAGI S. Discriminative methods for multi - labeled classification [J]. Advances in Knowledge Discovery and Data Mining,2004,23(4):22-30.

[27] TSOUMAKAS G, KATAKIS I. Multi-label classification; an overview[J]. International Journal of Data Warehousing and Mining,2006,3(3):34-56.

[28] HANG Minling, ZHOU Zhihua. A review on multi - label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering,2014,26(8):1819-1837.

[29] 王 进,王 鸿,夏翠萍,等. 基于 Spark 的组合分类器链多标签分类方法[J]. 中国科学技术大学学报,2017,47(4):350-357.

+++++

(上接第 36 页)

recognition via sparse spatio-temporal features[C]//Joint IEEE international workshop on visual surveillance and performance evaluation of tracking and surveillance. Beijing, China; IEEE,2006;65-72.

[15] 王 博,赵君喜. 基于时空兴趣点的人体行为识别方法研究[D]. 南京:南京邮电大学,2014.

[16] SCOVANNER P, ALI S, SHAH M. A 3-dimensional sift descriptor and its application to action recognition[C]//Proceedings of the 15th international conference on multimedia. Augsburg, Germany; ACM,2007;357-360.

[17] JI Xiaofei, WU Qianqian, JU Zhaojie, et al. Study of human action recognition based on improved spatio-temporal features[J]. International Journal of Automation and Computing,2014,11(5):500-509.

[18] 张世超,陈恩庆. 基于 Kinect 深度图像的动作识别[D]. 郑州:郑州大学,2016.

[19] HAERING N, VENETIANER P, LIPTON A. The evolution of video surveillance; an overview[J]. Machine Vision and Applications,2008,19(5-6):279-290.

[20] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE computer society conference on computer vision and pattern recognition. San Diego, CA, USA; IEEE,2005;886-893.

[21] 姬晓飞,左鑫孟. 基于关键帧特征库统计特征的双人交互行为识别[J]. 计算机应用,2016,36(8):2287-2291.

[22] 姬晓飞,周 路,李一波. 基于 AdaBoost 算法特征提取的人体动作识别方法[J]. 沈阳航空航天大学学报,2014,31(2):65-69.

[23] JI Yanli, YE Guo, CHENG Hong. Interactive body part contrast mining for human interaction recognition[C]//IEEE international conference on multimedia and expo workshops. Chengdu, China; IEEE,2014;1-6.

[24] SONG Sijie, LAN Cuiling, XING Junliang, et al. An end-to-end spatio-temporal attention model for human action recognition from skeleton data [J]. IEEE Intelligent Systems, 2016,31(2):45-53.

[25] LIN Liang, WANG Keze, ZUO Wangmeng, et al. A deep structured model with radius - margin bound for 3D human activity recognition [J]. International Journal of Computer Vision,2016,118(2):256-273.