doi:10.3969/j.issn.1673-4785.2011.05.008

电影中吸烟活动识别

叶果,程洪,赵洋

(电子科技大学自动化工程学院,四川成都611731)

摘 要:电影中的活动识别是计算机视觉领域的一个难点问题.传统识别算法受到电影中镜头视角变化、场景变化和光照变化等因素的影响,使得其对于真实场景活动识别的效果较差.针对上述问题,提出一种新颖的基于互信息的组合识别方法.该方法以纯贝叶斯互信息最大化构造初始框架,针对"吸烟"这类极具代表性的动作,将活动的SIFT信息和 STIP信息融合得到最优的组合分类器.该方法在电影《咖啡和烟》中进行了测试,实验结果表明,该方法具有很好的鲁棒性,并且很大程度上提高了抽烟活动的识别率.

关键词:电影;吸烟活动识别;纯贝叶斯互信息最大化;计算机视觉;模式识别

中图分类号:TP391.4 文献标志码:A 文章编号:16734785(2011)05-0440-05

Smoking recognition in movies

YE Guo, CHENG Hong,ZHAO Yang

(School of Automation, University of Electronic Science and Technology of China,Chengdu 611731,China)

AbstractAction recognition in movies is a difficult problem in the computer vision domain. Traditional approaches have a bad recognition effect because they are subjected to viewpoint changes,scene changes, and illumination changes in real scenes. This paper presented a novel combined recognition approach, using mutual information to solve the problems mentioned above. This method builds the initial skeleton using naive-Bayesian mutual information maximization(NBMIM)) and combines the shape information with the motion information to recognize smoking, which is a typical activity in movies. The proposed smoking recognition approach was evaluated in the film Coffee and Cigaretes. The results indicate that the proposed method is robust,and it significantly improves the recognition rate.

Keywords: movies; smoking action recognition; naive-Bayesian mutual information maximization;computer vision; pattern recognition.

人活动识别是计算机视觉与模式识别的重要研究领域,具有重要的学术价值和应用前景.自动活动识别和检索是数字媒体中的重要研究内容".传统的活动识别多集中在受限制环境中,而真实环境中的人体活动分析,由于电影、视频中有人的外形、动作、姿势变化、镜头远近变化、视角变化、周围环境变化的影响,导致人活动识别是一个极具挑战的问题13.为了减小这些因素的影响,先前人们的工作使用了大量简化的措施,比如限制镜头的运动,采用特定的固定场景,限定视角的变化等.

最近,广电总局办公厅发出《广电总局办公厅关于严格控制电影、电视剧中吸烟镜头的通知》,通知中指出,鉴于电影和电视剧在社会公众中的广泛影响,国家有关部门、社会各界要求严格控制电影和电视剧中吸烟镜头的呼声越来越强烈,电影和电视剧中过多的吸烟镜头,不符合我国政府控烟的基本立场,客观上有误导吸烟之嫌,容易对社会公众,特别是对未成年人产生不良影响.为避免电影和电视剧中个别镜头误导社会公众吸烟,特别是让未成年人远离烟草,倡导健康生活方式,培育社会文明,要进一步控制电影和电视剧中的吸烟活动在整个电影长度中的比例.对每年大量的电影、视频进行抽烟活动长度的人工统计几乎不可能.如何快速自动地对

收稿日期:2011-03-29.
基金项目:国家"973"计划资助项目(2011CB707000);国家自然科学基金资助项目(61075045);中央高校基本科研业务费专项基金资助项目(ZYGX2009X013).
通信作者:叶果.E-mail: yeguo0112@gmail.com.

电影中的吸烟活动进行识别并统计其在整个电影中的比例是解决上述问题的关键.

此外,据中国疾病预防控制中心统计,从2002—2010 年,中国烟民的数量仍在 3 亿以上,居高不下.吸烟已经深深危害了人们的身体健康.自动检测公共场所以及特定人群的吸烟活动具有重要意义,这将为社会调查和电子健康评估提供一种便利.

为了解决上述问题,本文提出了一种基于时空兴趣点(spatio-temporal interest point, STIP)[4]和尺度不变特征变换(scale invariant feature transform, SIFT)[5]的纯贝叶斯互信息最大化组合分类器(naive-Bayesian mutual information maximization, NB-MIM)进行吸烟活动识别.这种分类器不仅能识别出"衔烟"这样的静态行为,而且具有很好的鲁棒性,该方法很大程度上提高了电影中抽烟活动的正确识别率.

1　相关工作与意义

活动识别与分析如今在计算机视觉中是一个很热门的研究方向,并且已经出现了很多解决其中问题的方法.其中一种方法是通过运动轨迹来判断活动,这需要特定的目标跟踪[6-7],还有一种是通过身体的轮廓来判断人的活动,这需要去掉其背景[8].很多新方法都在不断地被发掘出来[9-10],现在的分类方法大多是通过局部时空特征来判定,并且使用的特征是形状和运动信息[11-13].

最近也有与本文相近的一些研究,袁浚菘提出一种纯贝叶斯互信息最大化的方法在活动识别方面取得了比较好的效果[14];一种自动提取电影中各种片段的方法最近也被提出[2];Laptev 提出了一种基于关键帧的方法在电影的人活动识别中得到了应用[15],并且他在实验中对于各种活动分类方法在电影中的识别效果进行了比较[1].但是以上这些方法只是对于点烟、吸烟这样的动态动作的检测,却无法识别"衔烟"这样的静态行为[15],或者只能基于图片识别"衔烟"这样的静态行为[16].而往往这 2 种动作会同时交替出现在电影中,现有的方法并不能很好地识别出来,Laptev 提出的最新组合算法在对电影的 8 种常见动作进行分类,其中最好的一类正确识别率为 53.3%.目前并没有一种专门针对于电影中吸烟活动的分类器,而这也是本文的意义所在.

2　识别系统框架

本文的吸烟活动检测系统包含了 2 个步骤:训练和识别,如图 1 所示.与传统的识别方法不同,本文将识别电影中所有的吸烟活动.传统的方法只关注点烟、吸烟这样短暂的动态动作,但是电影中会存在大量的"衔烟"这样的静态场景,加上电影中经常出现镜头切换、视角变化、光照变换、其他人活动等多种因素的影响,只使用运动特性识别,其效果较差.本文将同时基于形状信息 STFT 和运动信息 STIP,使用纯贝叶斯互信息最大化组合分类器来进行识别.与添加关键帧的方法[15]相比,本方法不需要大量的人工标注和提取关键帧,而是自动计算每一帧,从而能够快速自动地检测和识别吸烟行为.

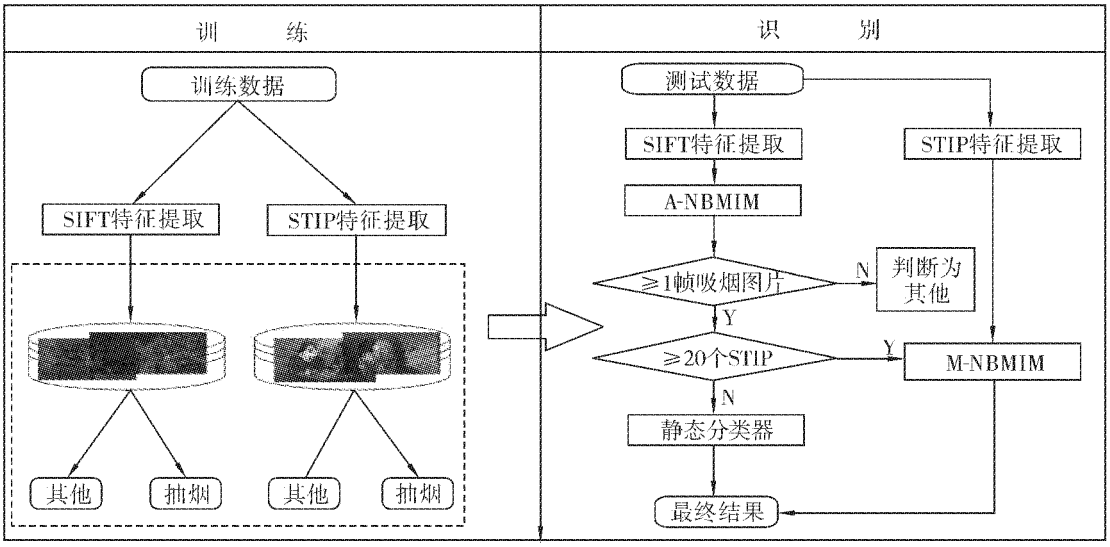


图 1　识别系统框架

Fig. 1　The recognition system framework

本文吸烟活动识别算法流程如下:图 1 左边是本文的训练模块,首先,提取视频每一帧的 SIFT 特征点和视频段的时空兴趣点 STIP,然后,从训练数据中生成特征池,这些特征描述符按其活动类别分为"吸烟"活动与"其他"活动,最后基于前述的训练模块,进行图 1 右子图的识别.在提取完视频的特征后,对于测

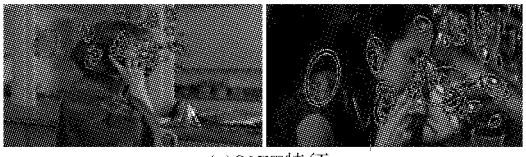
试视频的吸烟活动的检测识别,主要分3步来实现:1)使用SIFT信息和外形-纯贝叶斯互信息最大化分类器(A-NBMIM)对视频段的每一帧进行分类.如果判断出视频段中含1帧以上的吸烟图片,就将其视频段保留;否则,认为该视频段为"其他"类.通过这一步可以降低后续步骤分类对"其他"活动分类错误的概率.2)使用STIP特征信息和运动-纯贝叶斯互信息最大化分类器(M-NBMIM)对提取的视频段进行分类.考虑到在测试样本的特征点数过少时进行分类会出现偶然性误差较大的情况和计算的分数相同不能判断的情况,这里点数少于20个的视频段将不予以计算,从而在这一步,将会分出"吸烟"、"其他"、"不能判断"这3类动作.3)针对第2步中出现的"不能判断",根据前面使用SIFT信息和A-NBMIM的分类结果,统计吸烟帧数占视频段的比例,若大于50%,就将这一段视频定义为"吸烟",反之定义为"其他".按照以上算法完成对所有测试样本的识别并统计出电影中的抽烟活动.

3 吸烟活动识别算法

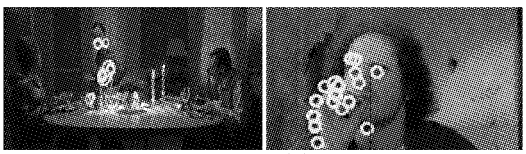
3.1 特征描述

3.1.1 SIFT特征

SIFT特征是图像的局部不变特征,它非常适合对不同图像或场景中的同一目标进行匹配,具有很高的鲁棒性.它对图像的光线亮度变化、尺度缩放以及旋转都能保持不变,对视角变化和噪声的出现也保持一定程度的稳定性,适用于海量数据库中进行快速的实时匹配,在目标识别中取得了良好的应用.图2中上面2幅图为活动分析中基于视频帧提取的SIFT点.



(a)SIFT特征



(b)STIP特征

图2 活动识别中的SIFT和STIP点

Fig. 2 The SIFT and STIP points in action recognition

提取SIFT特征步骤如下[5].

1)检测潜在兴趣点及其尺度:首先建立图像金字塔,然后利用高斯微分(difference-of-Gaussian,DOG)识别对尺度和方向不变的潜在兴趣点.

2)检测兴趣点:对上述产生的潜在兴趣点,根据稳定性度量选择稳定的兴趣点.

3)赋予兴趣点主方向:利用兴趣点邻域像素的梯度方向直方图计算每个关键点的主方向,使兴趣点的描述具有旋转不变性.值得注意的是,一个兴趣点可能存在多个主方向,这在实际使用中提高了局部描述器的鲁棒性.

4)生成特征点描述矢量:根据前面得到的兴趣点的位置、最优尺度以及主方向,将该图像块划分成4×4的子块,每个子块量化成8个方向.因此,得到一个128维的局部特征描述器,并将其归一化成2-范数为1的矢量,量化后的局部描述器具有亮度不变性.

3.1.2 STIP特征

本文把动作表示成一个时空目标,并且用一个时空兴趣点集(STIPs)[4]来描述它.与在二维图像中用到的SIFT特征不同,STIP特征是对三维视频中不变特征的扩展.提取完STIP特征之后,可以用以下2类特征来描述它们[1]:梯度直方图(histogram of oriented gradient, HOG)和光流直方图(histogram of flow, HOF).其中梯度直方图是一个72维的矢量,描述的是外形特征;光流直方图是一个90维的矢量,描述的是局部运动特征.由于STIP特征对于三维视频来说是局部不变的,所以这种特征对于动作变化相对鲁棒,而这种变化往往是由于动作的速度、尺度、光照和衣服等引起的.图2中下面2幅图为活动分析中基于视频序列提取的STIP特征点.

3.2 NBMIM方法

纯贝叶斯互信息最大化(naive-Bayesian mutual information maximization,NBMIM)的方法[14]在活动识别方面取得了比较好的效果.在本文系统中将采用此种方法结合各种特征来形成组合分类器.

本文用时空目标来表现动作,提取视频序列的STIP特征,用$V=\{I_t\}$表示一个视频序列,其中每一帧I_t由收集的STIPs构成,那么$I_t=\{d_i\}$.然后用$Q=\{d_i\}$表示一个视频段的STIP,$C=\{1,2,\cdots,c,\cdots,N\}$代表种类的标记集合.

基于纯贝叶斯假设和每个STIP间相互独立的假设可以得到一个视频段Q与一个特定类别$c\in C$的互信息为

$$\mathrm{MI}(C=c,Q)=\sum_{d_q\in Q}S^c(d_q).$$

先假定各类别出现概率相等,也就是$P(C=c)=\frac{1}{N}$,$S^c(d_q)$可以由式(1)表示:

$$S^c(d_q)=\log\frac{N}{1+\dfrac{p(d_q\mid C\neq c)}{p(d_q\mid C=c)}(N-1)}. \quad (1)$$

通过高斯核与最近邻近似得到其中的似然率如式(2).

$$\frac{p(d_q\mid C\neq c)}{p(d_q\mid C=c)}\approx \lambda^c\exp\left\{-\frac{1}{2\sigma^2}(\|d-d_{NN}^{c-}\|^2-\|d-d_{NN}^{c+}\|^2)\right\}. \quad (2)$$

式中:$\lambda^c = \frac{|T^{c+}|}{|T^{c-}|}$. $T^{c+} = \{V_i\}$ 代表 c 类的正训练样本,$V_i \in T^{c+}$ 是 c 类样本的一个视频段,这里将正训练样本的 STIPs 数据表示为 $T^{c+} = \{d_j\}$;同样,负样本可以表示为 T^{c-},其包含所有的负 STIPs.

因此,对于每一个与 c 类相关的 STIP,这里调整其分数为

$$S^c(d) = \log \frac{N}{1 + \lambda^c e^{[-r(d)\omega_\sigma(d)]}(N-1)}.$$

式中:$r(d) = \| d - d_{NN}^{c-} \|^2 - \| d - d_{NN}^{c+} \|^2$,并且 $\omega_\sigma(d) = \frac{1}{2\sigma^2}$是 d 周围训练样本的纯度.

最后,本文可以通过计算每一视频段中 STIP 或者 SIFT 点的得分来判断其属于哪一类.

4　实验结果及分析

本文使用文献[1,15]中提供的视频段以及在《风声》等电影中截取的吸烟片段,这些活动就出现在不同场景,被不同的人表现出来,并且从不同的角度被拍摄记录. 然后分别提取其视频段的 STIP 和每一帧的 SIFT 特征点.

4.1　训练数据

对于“吸烟”活动,采用的《风声》、《热血高校》、《革命之路》电影中的吸烟片段作为训练样本,共 110 个小视频段. 然后提取其 STIP 点特征,共89 908个点,同时,为避免大量重复的特征带来的计算浪费,每隔 25 帧提取其 SIFT 特征,共37 687个点.

对于“其他”活动,采用《阿甘正传》、《蝴蝶效应》等电影中的片段,包含站立、坐下、握手、拥抱、坐起、打电话、走出车、接吻等多种主要动作以及其他杂乱动作[1]. 使用了 12 个大视频片段,使提取点数相接近,以避免训练样本不均匀带来的影响,共提取了86 810个 STIP 点,以及38 326个 SIFT 特征点.

4.2　测试数据

对于测试数据,使用电影《咖啡和烟》中的吸烟片段[15]以及《低俗小说》、《火星任务》等电影的非吸烟活动片段[1]. 由于电影《咖啡和烟》按场景与主题分为了 11 个片段,相当于从 11 部电影中提取数据,所以就只在这些数据中进行实验. 测试数据共 84 个视频段,包含 42 个吸烟样本和 42 个其他活动样本,分别提取其每一帧的 SIFT 特征点和视频段的 STIP 特征.

这些训练数据和测试数据使用的电影不相同,所以均没有主题和背景上的重叠. 训练与测试样本示例如图 3 所示,图中上 2 行对应的为“吸烟”动作,下 2 行对应的为“非吸烟”动作,即 “其他”类. 详细视频数据见网址 www. uestcrobot. net/smokings.

(a)训练样本　　(b)测试样本

图 3　实验中使用的训练样本与测试样本

Fig. 3　Examples of training samples and testing samples in our experiments

4.3　实验结果及分析

本文实验中参数为经验参数 $\lambda = 1, \sigma = 2.6$,可以达到最优的实验效果,分别使用文献[4-5]提供的方法来提取 STIP 特征和 SIFT 特征.

使用 M-NBMIM 方法进行分类实验,其结果如表 1 所示. 可以看出,只使用 STIP 特征,系统的识别率并不高. 原因在于 STIP 不包含静态外形信息,所以“衔烟”这样的活动 STIP 点较少,或者没有特征点的动作不能判断,并且会有较多的“其他”类被误判为“吸烟”类,从而总体识别率不高.

表 1　基于 M-NBMIM 的实验结果

Table 1　The results based on M-NBMIM

标记类别	实验结果			
	吸烟	其他	点少于 20	正确率/%
吸烟	24	11	7	57.1
其他	8	34	0	81.0

下面将按照本文提出的纯贝叶斯互信息最大化组合分类器进行分类. 先用 SIFT 信息找出不含吸烟片段的视频段,将其定义为“其他”. 然后对剩下的视频段使用 STIP 信息进行初步分类. 由于吸烟这个动作的时间短暂性和拥有大量“衔烟”这样的静态行为,因此出现了许多点数过少的情况. 这种情况用于计算会导致大量偶然因素,使得结果不能真实稳定,所以实验中直接将此种情况提取出来,不使用 STIP 点计算分类. 然后再使用点数少于 20 的视频段包含的所有帧的 SIFT 特征进行计算,得到最终分类结果,结果如表 2 所示.

表 2　基于提出的组合分类器的实验结果

Table 2　The results based on combined classifier

标记类别	实验结果		
	吸烟	其他	正确率/%
吸烟	29	13	69.0
其他	5	37	88.1

从表 2 结果可以看出,在使用了 SIFT 特征后,在第 1 步中,能够将“其他”这类动作的识别错误率

降低,在第3步中,能提高吸烟这种特定动作的识别正确率,这样系统对于吸烟活动的识别率也就得到了大幅度的提升.

5 结束语

本文主要对真实电影中的人的抽烟行为进行识别,与之前在特定场景中分析人的活动相比,这里是在包括人物外表改变、场景变换、镜头视角变换和动作时间改变的真实场景中进行活动分析与识别.在真实场景的识别活动中,由于各种因素的影响,导致现在很多在特定视频中识别效果比较好的方法在真实电影中的识别效果很低.考虑到若只使用单独运动信息或形状信息在真实场景中识别效果不高,因此采用了一种纯贝叶斯互信息最大化组合分类器作为统一的计算框架,实验结果证明此方法相比于传统方法提高了识别率.

但是,使用视频中帧的信息的方法,对于包含物品的运动比较有效,如吸烟、喝水,而对于诸如走路、慢跑、跑步这样动作相似的行为识别效果一般.如何将这种方法运用到其他所有动作以及如何减少运算时间都将是今后研究的重点方向.

参 考 文 献 :

[1] LAPTEV I,MARSZALEK M,SCHMID C,et al. Learning realistic human actions from movies [C]//Proceedings of CVPR: IEEE Conference on Computer Vision and Pattern Recognition.Anchorage,USA,2008: 1-8.

[2] GAIDON A,MARSZALEK M,SCHMID C. Mining visual actions from movies[C]//Proceedings of BMVC: British Machine Vision Conference. London,UK,2009: 1-11.

[3] WANG JZ,GEMAN D,LUO Jiebo,et al. Real-world image annotation and retrieval: an introduction to the special section[J].IEEE Transactions on Patten Analysis and Machine Inelligence,2008,30(11): 1873-1876.

4] LAPTEV I. On space-time interest points[J]. Interational Journal of Computer Vision,2005,64(2/3) : 107-123.

[5] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004,60(2) ; 91-110.

[6] ALI S,BASHARAT A, SHAH M. Chaotic invariants for human action recognition[C]//Proceedings of ICCV: IEEE International Conference on Computer Vision. Rio de Janeiro,Brazil,2007: 14-21.

[7] NGUYEN N T,PHUNG D Q,VENKATESH S,et al. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov models [C] //Proceedings of CVPR: IEEE Conference on Computer Vision and Pattern Recognition. San Diego,USA,2005: 955-960.

[8] MOESLUND T B,HILTON A,KRUGER V. A survey of advances in vision-based human motion capture and analysis [J] . Computer Vision and Image Understanding,2006, 104(2)) ; 90-126.

[9] DUAN Liin,XU Dong,TSANG IW,et al. Visual event ecognition in videos by learring from web data[C]//Proceedings of CVPR: IEEE Conference on Computer Vision and Patterm Recognition. San Francisco,USA,2010: 1959-1966.

10] CAO Liangliang,LU Zicheng,HUANG T. Cros-dataset action detection[C] //Proceedings of CVPR: IEEE Conference on Computer Vision and Pattern Recognition. San Francisco,USA,2010: 1998-2005.

[11] NATARAJAN P,NEVATIA R. View and scale invariant action recognition using multiview shape-flow models [C] //Proceedings of CVPR: IE Conference on Computer Vision and Pattern Recognition. Anchorage,USA, 2008: 1-8.

[12] VITALADEVUNI S N,KELOKUMPU V,DAVIS L s. Action recognition using ballistic dynamics[C]//Proceedings of CVPR: IEEE Conference on Computer Vision and Patten Recognition.Anchorage,USA,2008:1-8.

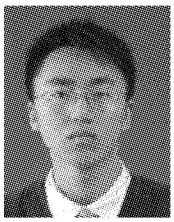
[13] YILMAZ A,SHAH M. Actions sketch: a novel action representation[C]//Proceedings of CVPR: IEEE Conference on Computer Vision and Patten Recognition. San Diego, USA,2005: 984-989.

[14] YUAN Junsong,LIU Zicheng,WU Ying. Discriminative subvolume search for efficient action detection[C]//Proceedings of CVPR: IEEE Conference on Computer Vision and Pattern Recognition. Miami,USA,2009: 2442-2449.

15] LAPTEVI,PEREZ P. Retrieving actions in movies[C] // Proceedings of ICCV: IEEE International Conference on Computer Vision. Rio de Janeiro,Brazil,2007: 1-8.

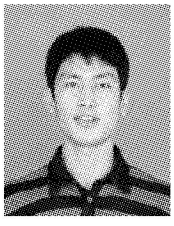
[16] WU Pin,HSIEH JH,CHENG J C,et al. Human smoking event detection using visual interaction clues[C]//Proceedings of ICPR: IEEE International Conference on Pattern Recognition. Istanbul,Turkey,2010:4334-4347.

作者简介:



叶果,男,1990年生,本科生,主要研究方向为人的活动识别、计算机视觉与模式识别.

程洪,男,1973年生,教授,博士生导师,博士,EEE和 ACM会员,2006-2009年在美国卡内基-梅隆大学计算机学院进行博士后研究.主要研究方向为机器人、计算机视觉与模式识别、机器学习.先后主持和参与包括国家"973"计划项目、国家"863"计划项目,以及重要企业横向项目等10余项科研项目.发表学术论文40余篇,出版教材和专著各1部.



赵洋,男,1988年生,硕士研究生,主要研究方向为计算机视觉与模式识别