DOI:10.11721/cqnuj20140519

基于粒子滤波的行为识别方法*

周 卫 春
(绵阳职业技术学院 人文科学系, 四川 绵阳 621000)

摘要:行为识别是图像处理的一个热点问题。一些研究表明在监督学习的框架下,通过提取时空兴趣点(Spatial-temporal interest points)能较好地识别人体的行为。由于兴趣点中包含与人体行为无关的噪声点,为了改进兴趣点的提取,提出了一种基于人体骨架的改进方法。该方法通过粒子滤波(Particle filter)算法改进人体骨架的精度,而改进后的人体骨架,能得到更有效的兴趣点。通过在"Weizmann","KTH"数据集的测试,实验结果表明,该算法不仅能够提高人体行为的识别,而且能够改进人体骨架的精确度。

关键词:行为识别;时空兴趣点;粒子滤波;姿态估计;支持向量机
中图分类号:TP391.41　　**文献标志码**:A　　**文章编号**:1672-6693(2014)05-0105-05

行为识别是图像处理的一个热点问题。行为识别算法可以分为两大类:1)基于时空兴趣点的方法[1-2]。该方法将人体运动过程表示为一系列与时间空间相关的点,通过 SVM(支持向量机)进行分类;2)基于人体模板。该方法通过识别出人体运动部位,找出相关部位的行为模式[3-4]。由于人体模板本身存在精确度不高的问题,而且这类方法主要采用时间序列分析,而人体行为模式是很典型的非线性,非高斯分布的时间序列,所以本文采用的是第一类方法。

基于时空兴趣点的方法,将人体一系列的行为表示成词袋(Bag-of-word)。兴趣点的计算是基于 Gabor 滤波。兴趣点的精确度由 Gabor 滤波器的阈值决定。兴趣点是分布在人体四肢、躯干的一系列点。一种直观的方法就是将远离人体骨架的兴趣点作为噪声(异常点)处理。由于以前人体骨架提取的精确度不高,较少有研究将人体骨架和兴趣点结合起来,用于行为的识别。

人体骨架的提取是姿态估计(Pose estimation)的主要研究问题。目前的研究主要是提取静态图片中人体的骨架[5-7]。本文研究的对象是动态运动视频中的人体行为。粒子滤波是跟踪人体行为的一种方法。现有的研究主要侧重在跟踪,较少有研究者将粒子滤波应用到行为识别。本文通过粒子滤波预测人体骨架的在运动中的可能位置,缩小搜索的范围,提高人体骨架的精确度,从而提高兴趣点的精度。

本文的主要贡献是提出一种通过人体骨架改进兴趣点的方法,该方法在改进兴趣点的同时,能够提高运动中人体骨架提取的精确度。而更精确的人体骨架能够得到更有效的兴趣点,最终达到提高行为识别准确率的目的。

1 基于人体骨架的兴趣点

本节首先给出行为识别的方法流程图(图 1),然后介绍了基于 Gabor 滤波器的兴趣点和人体骨架提取的算法。最后提出一种基于人体骨架改进兴趣点的方法。行为识别的流程描述如下:

1)输入视频,如果是第一帧,进行初始化操作,主要是提取人体的骨架。

2)如果不是第一帧,则基于粒子滤波跟踪算法和前后帧图像的变化,预测运动中人体骨架的可能位置。

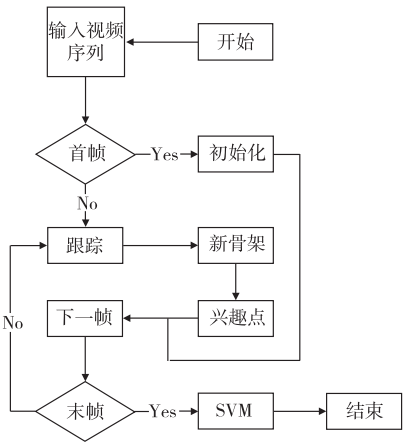


图 1　方法流程

* 收稿日期:2014-05-06　　修回日期:2014-06-05　　网络出版时间:2014-9-17 22:37
作者简介:周卫春,女, 讲师,研究方向为高等数学、工程数学、图像处理,E-mail:449076939@qq.com
网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20140917.2237.020.html

3)计算出新的人体骨架。

4)基于人体骨架和前后帧图像的变化,计算出兴趣点的分布。

5)读取下一帧,如果不是最后一帧,则返回2)。

6)如果是最后一帧,则将产生的兴趣点序列输入训练好的SVM分类进行行为识别。

1.1 基于Gabor滤波器的兴趣点

本文中的兴趣点在于采用文献[8]中提出的Gabor滤波器:

$$R=(I*g*h_{ev})+(I*g*h_{od}) \tag{1}$$

其中,I是输入的图像,$g(x,y;\sigma)$是高斯核,h_{ev}和h_{od}是一维Gabor滤波器,其定义如下:

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

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

通过设置阈值,当(1)式中的R大于该阈值,得到相应的兴趣点。这种方法会产生大量与行为无关的兴趣点。

1.2 姿态估计

人体骨架的提取是姿态估计的主要研究问题。本文采用文献[7]中提出的算法。人体骨架各部分相似度的计算如下:

$$S(t)=\sum_{i\in V}b_i^{t_i}+\sum_{ij\in E}b_{ij}^{t_i,t_j} \tag{3}$$

其中$p_i=(x,y)$表示人体骨架的第i部分,比如上肢。t_i表示第i部分的运动类型,比如向前、向后运动。该算法假设人体骨架为树形结构$G(V,E)$,V表示节点,E表示边。$b_i^{t_i}$表示第i部分,运动类型为t_i的值。$b_{ij}^{t_i,t_j}$表示第i部分运动类型为t_i,第j部分运动类型为t_j的值。

人体骨架各部分的位置,及其相关运动模式的计算如下:

$$S(I,p,t)=S(t)+\sum_{i\in V}w_i^{t_i}\cdot\Phi(I,p_i)+\sum_{ij\in E}w_{ij}^{t_i,t_j}\cdot\psi(p_i-p_j) \tag{4}$$

其中,$\Phi(I,p_i)$表示p_i在图像I中的特征向量。$\psi(p_i-p_j)$表示第i部分与第j部分之间的距离。$w_i^{t_i}$计算第i部分运动类型为t_i的权值。$w_{ij}^{t_i,t_j}$计算第i部分运动类型为t_i,第j部分运动类型为t_j的权值。

1.3 基于人体骨架的兴趣点

本节基于1.1和1.2节方法,提出一种改进兴趣点,去掉与人体行为无关的兴趣点的方法。基本思想如图2所示。图2(b)是前后帧差分图,反应图像的变化(为了便于说明,简化一些复杂情况,图2采用的是原理图)。图2(b)中虚线方框以外的黑色区域表示非人体运动变化的区域。虚线方框中的黑色区域与人体运动有关,记为A_d^t(上标t表示当前为第t帧)。图2(a)是基于1.2节方法生成的人体骨架区域,每个部分是一个矩形的区域。该区域记为A_s^t(上标t表示当前为第t帧)。所以与人体运动相关的区域为

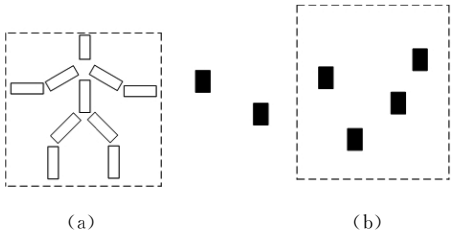


图2 基于人体骨架的兴趣点

$$A^t=A_d^t\cup A_s^t\cup A_s^{t-1} \tag{5}$$

即兴趣点应该落在前后两帧的人体骨架和其相关的差分区域中。

该方法的最大好处是不用设置固定的阈值。先设置阈值,使生成的兴趣点个数为0,然后减小阈值。ΔN_A表示减小阈值后,A^t区域中兴趣点增加的个数。基于实验,本节提出一种启发式的减小阈值的方法,当连续3次减小阈值,相应的ΔN_A都在减小,则结束。

2 跟踪算法

1.2节人体骨架的提取算法的最大缺点是,人体骨架的精度受搜索图像范围的影响。当图像中包含越多与人体骨架无关的内容时,精度越低。本节基于粒子滤波的人体跟踪,提出一种有效缩小人体骨架搜索范围的

方法。

2.1 粒子滤波

在状态空间模型中，隐性状态$\{x_t;t\in N\}$，可以通过马尔科夫过程进行分析。设初始分布为 $p(x_0)$，可观察的状态为$\{y_t;t\in N^*\}$，粒子滤波算法，递归地计算后验概率 $p(x_t|y_{1:t})$。其核心思想是基于第 T 帧之前的观察值推测出第 T 帧的隐性状态。递归公式[10]如下：

预测：
$$p(x_t\mid y_{1:t-1})=\int p(x_t\mid x_{t-1})p(x_{t-1}\mid y_{1:t-1})\mathrm{d}x_{t-1} \tag{6}$$

更新：
$$p(x_t|y_{1:t})=\frac{p(y_t|x_t)p(x_{t-1}|y_{1:t-1})}{\int p(y_t|x_t)p(x_t|y_{1:t-1})\mathrm{d}x_t} \tag{7}$$

其中，$p(x_t|x_{t-1})$表示隐性状态从第 $T-1$ 帧变到第 T 帧的概率。$p(y_t|x_t)$表示第 T 帧隐性状态下第 T 帧观察值 y_t的概率。(6)式表示基于 $T-1$ 帧之前的观察值预测第 T 帧的隐性状态 x_t。(7)式表示根据第 1～T 帧的观察值计算第 T 帧隐性状态的概率。(6)、(7)式中的积分可以通过采样的方法进行计算。后验概率 $p(x_{0:t}\mid y_{1:t})$可以近似计算：

$$\hat{P}_N(\mathrm{d}x_{0:t}\mid y_{1:t})=\sum_{i=1}^{N}\hat{\omega}_t^{(i)}\delta_{x_{0:t}^{(i)}}(\mathrm{d}x_{0:t}) \tag{8}$$

$$\hat{\omega}_t^{(i)}\propto\hat{\omega}_{t-1}^{(i)}p(y_t\mid x_t^{(i)}) \tag{9}$$

(8)式中 $\delta_{x^0}(x)$为狄拉克函数。(9)式中 $\hat{\omega}_t^{(i)}$ 为归一化的重要性权值。关于(6)～(9)式详细的讨论，可参考文献[10]。

2.2 基于粒子滤波的人体跟踪

本文采用文献[9]中基于粒子滤波的跟踪方法。$T+1$ 帧的跟踪目标的计算如下：

$$x_{t+1}=Ax_t+Bx_{t-1}+Cv_t,v_t\sim N(0,\sum) \tag{10}$$

其中，$x_t=(d_t,d_{t-1},s_t,s_{t-1})$，$d=(x,y)$ 表示追踪目标在图像中的位置，s 表示目标在图像中的比例，目标搜索面积为

$$R(x_t)=d_t+s_tW \tag{11}$$

其中，W 表示矩形窗口的大小，该区域的颜色分布的核密度估计为

$$q_t(n;X)=K\sum_{u\in R(x)}w(|u-d|)\delta[b_t(d)-n] \tag{12}$$

其中，$b_t(u)$表示第 t 帧图像中，位置 u 处的颜色向量。K 是规范化常量。

$$p(y_t\mid x_t)\propto\exp\left\{-\lambda\left[1-\sum_{i=1}^{N}\sqrt{q^*(n)q_t(n;x)}\right]\right\} \tag{13}$$

其中，$q^*(n)$是基于(12)式计算的目标参考分布。基于 $p(y_t\mid x_t)$ 利用粒子滤波实现人体跟踪。关于粒子滤波的具体内容，可参考文献[10]。

2.3 改进的跟踪搜索区域

2.2 节对跟踪目标的预测是基于(10)式二阶自回归方程。本节前后帧所反映出的运动变化，给出跟踪目标的预测区域，并基于这个新的预测区域计算人体骨架。图 3(a)中的两个矩形区域分别表示(5)式的 A^{t-1}，A^{t-2}，ΔT^t 表示前 2 帧中矩形的上边差的绝对值；ΔB^t 表示前 2 帧中矩形的底边差的绝对值；ΔR^t 表示前 2 帧中矩形的右边差的绝对值；ΔL^t 表示前 2 帧中矩形的左边差的绝对值。

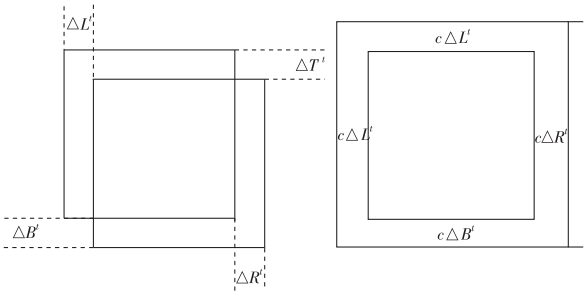


图 3　跟踪区域

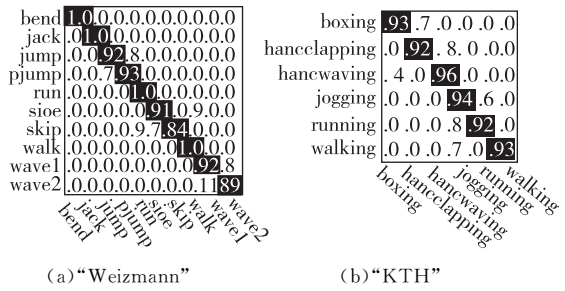
新的搜索面积 A_p^t 为图 3(b)中外部较大的矩形，图 3(b)中内部较小的矩形为 A^{t-1}。A_p^t 由 A^{t-1}向上下左右扩大 $c\Delta T^t$，$c\Delta B^t$，$c\Delta L^t$，$c\Delta R^t$(实验中 $c=3$)。

2.2 节中，(10)式的 x_{t+1}，(11)式中的 W，以及(11)、(12)式中的 d 由本方法替代。

3 实验结果及分析

本实验所需硬件环境:Intel i5 CPU, 内存 8 GB,软件环境:Windows 7,Opencv 2.43。本实验数据,来自"Weizmann"[11], "KTH"[12]。"Weizmann"包含 10 种行为类型,大约 100 段视频。"KTH"包含 6 种行为类似,大约 2 300 段视频。行为识别采用的是支撑向量机(SVM),Radial basic function kernel。

图 4 显示的是基于本文方法的混淆矩阵(Confusion matrix)。对于"KTH",识别率为93.3%;对于"Weizmann",识别率为 94.1%。表 1 给出了与其他方法的比较。对于"KTH",本方法比其他文献方法的识别率高,对于"Weizmann",本方法虽然不如文献[14],但是对于"KTH"的识别率比文献[14]高。



(a)"Weizmann" (b)"KTH"

图 4 混淆矩阵

表 1 与其他方法的比较 %

	Weizmann	KTH
本文	94.1	93.3
文献[16]	90.05	83.3
文献[15]	92.89	91.33
文献[14]	100	90.5
文献[8]	85.2	81.17

本方法不仅能达到较好的行为识别,由于对人体骨架搜索面积的优化,对人体骨架提取的精度也有所提高。表 2、3 分别给出对于"KTH","Weizmann"数据集,分别采用文献[7]和本文方法得到的人体骨架的精度。文献[7]对下臂和小腿的精度较低。而本文方法比文献[7]对人体骨架的识别精度有较大改进。

4 结束语

本文提出一种基于人体骨架,改进兴趣点的方法。实验表明本方法不仅能提高人体行为的识别率,而且能够改进运动中人体骨架提取的精确性。本文方法主要是对兴趣点的模式通过 SVM 进行分类,如何得到层次化、结构化的兴趣点,是下一步需要研究的问题。虽然人体骨架的识别精度还不是很高,表 2、3 表明下臂和小腿的识别率比较低,但是如果不基于兴趣点,仅仅分析人体骨架的运动模式是否能够得到更好的行为识别,同样也是需要进一步研究的问题。

表 2 Weizmann %

方法	躯干	头	上臂	下臂	大腿	小腿
本文	100	87	88.5	69.2	86.6	75.2
文献[7]	88.5	73.2	64.7	43.8	77	62.3

表 3 KTH %

方法	躯干	头	上臂	下臂	大腿	小腿
本文	98.2	94.2	87.3	67.3	89.4	73.5
文献[7]	91.5	81.4	72.5	50.7	82.7	65.3

参考文献:

[1] Chakraborty B, Holte M B, Moeslund T B, et al. A selective spatio-temporal interest point detector for human action recognition in complex scenes[C]//Proceedings of the IEEE international conference on computer vision. USA: IEEE, 2011: 1776-1783.

[2] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]//Proceedings of the IEEE international conference on computer vision. USA: IEEE, 2008: 1-8.

[3] Blank M, Irani M, Basri R. Actions as space-time shapes [C]//Proceeding of international conference on computer vision. USA: IEEE, 2005: 1395-1402.

[4] Ke Y, Sukthankar R, Hebert M, Efficient visual event detection using volumetric features[C]//IEEE international conference on computer vision. USA: IEEE, 2005: 166-173.

[5] Ferrari V, Marin J M, Zisserman A, Progressive search space reduction for human pose estimation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. USA: IEEE, 2008: 1-8.

[6] Ferrari V, Marin J M, Zisserman A. Pose search: retrieving

people using their pose[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. USA: IEEE, 2009: 1-8.

[7] Yang Y, Ramanan D. Articulated pose estimation with flexible mixtures-of-part[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. USA: IEEE, 2011: 1385-1392.

[8] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal feature[C]//Proceedings of VS-PETS. USA: IEEE, 2005: 65-72.

[9] Hess R, Fern A, Discriminatively trained particle filters for complex multi-object tracking [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. USA: IEEE, 2009: 240-247.

[10] De Freitas N, Doucet A, Gordon N . An introduction to sequential Monte Carlo methods[M]. SMC Practice: Springer Verlag.

[11] Blank M, Irani M, Basri R, Actions as space-time shapes [C]//Proceeding of International conference on computer vision. USA: IEEE, 2005: 1395-1402.

[12] Laptev I, On space-time interest points[J]. Int J Comput Vis, 64(2-3): 107-123.

[13] De Freitas N, Doucet A, Gordon N. An introduction to sequential Monte Carlo methods[M]. SMC Practice: Springer Verlag.

[14] Fathi A, Mori G. Action recognition by learning mid-level motion features[C]//Proceedings of the IEEE computer society conference on computer vision and pattern recognition. USA: IEEE, 2008: 240-247.

[15] Zhang Z, Hu Y, Chan S, et al. Motion context: a new representation for human action recognition[C]//Proceedings of European conference on computer vision. USA: IEEE, 2008: 817-829.

[16] Niebles J, Wang H, Fei F L. Unsupervised learning of human action categories using spatial- temporal words[J]. Int J Comput Vis, 79(3): 299-318.

Recognizing Human Actions by Particle Filter

ZHOU Weichun

(Department of Humanities, Sichuan Mianyang Vocational and Technical College, Mianyang Sichuan 621000, China)

Abstract: Action recognition is a hot research topic in image processing area. Some studies have shown that based on supervised learning, spatial-temporal interest points which are extracted from images can recognize human action. Since interest points contain some noises which are not related to human action, a method which is based on human skeleton is presented to refine interest points. This method can improve the precision of human skeleton by particle filter. The refined human skeleton is used to get better interest points. Based on "Weizmann","KTH" dataset, experiment results show that the method can improve the precision of human action recognition and human skeleton.

Key words: human action recognition; spatial-temporal interest points; particle filter; pose estimation; support vector machine

(责任编辑 游中胜)