

文章编号: 2095-4980(2016)03-0443-06

基于深度信息的人体运动识别方法

陈 力, 王俊凯, 张玉玺

(北京航空航天大学 电子信息工程学院, 北京 100191)

摘 要: 近年来, 可见光视频序列的人体运动识别研究已经取得了一定的进展。由于这些数据源容易受到目标颜色、光照强度和背景杂波的影响, 因此将深度信息应用于人体运动识别。本文首先采用了基于时空兴趣点的人体运动的局部表征方法, 分别实现了 Harris 时空兴趣点与基于 Gabor 滤波器的时空兴趣点(STIPs)检测方法在深度信息上的应用。然后对相应结果进行立方体描述并提取了深度立方体相似特征(DCSF)。最后利用基于时空码本的支持向量机(SVM)动作分类器完成对动作的分类。实验表明, 基于 Gabor 滤波器的检测方法在深度数据集上取得了更好的识别效果。

关键词: 分层传输运动分析; 运动识别; 时空兴趣点; 运动表征; SVM 分类

中图分类号: TN911.73

文献标识码: A

doi: 10.11805/TKYDA201603.0443

Human motion recognition method based on depth information

CHEN Li, WANG Junkai, ZHANG Yuxi

(School of Electronic and Information Engineering, Beihang University, Beijing 100191, China)

Abstract: In recent years, recognition of human motion on visible light video sequences has made some progress. Since the data sources are sensitive to target color, light intensity and background clutters, the depth information is applied to human motion recognition. In this paper, a local representation method of human movement based on Spatio-Temporal Interest Points(STIPs) is adopted, and the applications of Harris and Gabor filter detection methods on depth information are achieved. A novel Depth Cuboid Similarity Feature(DCSF) is built to describe the corresponding results. Finally, action classification is completed by Support Vector Machine(SVM) classifier based on spatio-temporal codebook. Experimental results demonstrate that detection method of Gabor filter obtains better recognition performance in depth datasets.

Key words: motion analysis; motion recognition; spatio-temporal interest points; motion characterization; Support Vector Machine classification

人体运动识别是计算机视觉领域一个非常活跃的研究方向。运动识别的主要研究方法是通过适当的预处理将目标从嘈杂的背景中提取出来, 在检测出的目标图像中选择恰当的表征方法来提取相应的信息表征目标的运动状态, 最后通过模板匹配识别出目标运动模型^[1]。本文在深度信息的基础上研究了有关人体运动识别的方法。

近年来, 基于时空兴趣点的动作表征和识别^[2]取得了一些突破性的进展。其中应用比较广泛的是 Laptev^[3], 将二维图像中的 Harris 角点检测技术扩展到三维时空领域, 可以从视频序列中提取出丰富的时空兴趣点。Dollar^[4]提出了基于 Gabor 滤波器的时空兴趣点检测算法, 这些算法可以从视频中快速有效地检测出丰富的兴趣点。本文在同一实验环境中对目前常用的兴趣点提取算法进行性能评估, 围绕着时空兴趣点的周围定义了深度立方体相似特征描述符(DCSF)来提取特征数据^[5]。最后利用所提取到的特征集合构建了分类器, 完成动作的分类。

1 基于深度信息的人体运动识别概述

人体运动识别主要包括图像和视频序列中运动人体的检测、跟踪, 提取运动特征来对人体行为进行表征, 从而可以运用特征数据来理解人体运动状态。在运动人体检测之前一般需要对图像或视频序列进行预处理, 滤除图

像或视频中的噪声,然后从目标的运动信息中提取目标的运动特征来表征目标的运动状态,最后构建分类器将训练样本序列和测试序列的特征进行匹配,判断出当前动作处于哪种模型。人体运动识别的大致流程如图 1 所示。

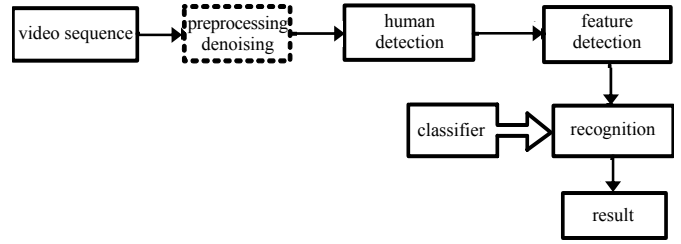


Fig.1 Flow chart of human motion recognition
图 1 人体运动识别的流程

2 时空兴趣点检测和特征提取

本文主要从人体运动的局部表征法出发,通过提取深度视频序列中的时空兴趣点(STIPs)来对运动的人体进行描述和表征。

2.1 Harris STIPs 检测

将二维空间域的角点检测的思想拓展到时空领域即可得到 STIPs 的概念。在深度视频序列中, STIPs 就是像素深度值同时在时域和空域发生剧烈变化的点。Laptev^[3]将 Harris 角点检测算法扩展到时空域中,得到 Harris STIPs 的检测方法。记深度图像序列为 $D(u,v,t)$ 。使用高斯核函数卷积运算将 $D(u,v,t)$ 映射到高斯空间,如式(1)所示,高斯核函数具体形式如式(2)所示^[3]:

$$L(u,v,t;\sigma_l^2,\tau_l^2) = g(u,v,t;\sigma_l^2,\tau_l^2) * D(u,v,t) \quad (1)$$

$$g(u,v,t;\sigma_l^2,\tau_l^2) = \frac{1}{\sqrt{(2\pi)^3 \sigma_l^4 \tau_l^2}} \exp\left\{-\left(u^2 + v^2\right)/2\sigma_l^2 - t^2/2\tau_l^2\right\} \quad (2)$$

将深度图像序列映射到高斯空间后,由于增加了时域信息,时空域中的二阶矩阵由空间域中的二阶矩阵变为一阶偏微分导数乘积构成的三阶方阵。对一阶微分偏导数的平滑仍然采用高斯窗口函数,时间和空间上的方差分别记为 σ_l^2 和 τ_l^2 ,如式(3)所示^[3]:

$$w(u,v,t;\sigma_l^2,\tau_l^2) = \frac{1}{\sqrt{(2\pi)^3 \sigma_l^4 \tau_l^2}} \exp\left\{-\left(u^2 + v^2\right)/2\sigma_l^2 - t^2/2\tau_l^2\right\} \quad (3)$$

最后由二维高斯窗口和三阶方阵卷积得到的观测矩阵 μ 如式(4)所示:

$$\mu = w(u,v,t;\sigma_l^2,\tau_l^2) * \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \quad (4)$$

式中:图像序列 $D(u,v,t)$ 和高斯核函数 $g(u,v,t;\sigma_l^2,\tau_l^2)$ 卷积运算结果沿着空间域 x,y 方向和沿着时间域 t 方向的一阶微分偏导数如式(5)所示:

$$\begin{cases} L_x(u,v,t;\sigma_l^2,\tau_l^2) = \partial_x(g * D) \\ L_y(u,v,t;\sigma_l^2,\tau_l^2) = \partial_y(g * D) \\ L_t(u,v,t;\sigma_l^2,\tau_l^2) = \partial_t(g * D) \end{cases} \quad (5)$$

式(5)的二阶矩阵 μ 与二维空间域中的二阶矩阵有相同的特性,通过比较 μ 的 3 个特征值 λ_1, λ_2 和 λ_3 的大小就可以检测出时空兴趣点的位置。将二维空间域中的响应函数 R^{sp} 扩展到三维时空域中,得到式(6):

$$R = \det(\mu) - k \times \text{trace}^3(\mu) = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3 \quad (6)$$

R 的正局部最大值所对应的点就是图像深度值在空间域和时间域上同时发生剧烈变化大的点。这样,通过搜索 R 的局部最大值,就可以得到深度图像序列的时空兴趣点。

2.2 基于 Gabor 滤波器的 STIPs 检测

Harris STIPs 检测算法具有良好的数学性能,能够较为有效地检测出 STIPs。但是在某些领域,图像序列中像素值在时间域上的变化往往不如空间域上的变化剧烈,导致检测时的响应强度较低,使得真正能得到的 Harris STIPs 十分稀少,为了解决上述存在的问题,Dollár 等人提出了一种新的 STIPs 检测算法,在空间域上仍使用高斯滤波器滤波,但在时间域上改为用一维 Gabor 滤波器。

与 Harris 检测方法在时空域均使用高斯滤波不同的地方在于该算法在时间域使用一维 Gabor 滤波器,在空间

域中仍然使用高斯滤波器。该算法整体的响应函数可以近似为式(7)的形式^[4]:

$$R_G(u, v, t) = (D * g * h_{ev})^2 + (D * g * h_{od})^2 \quad (7)$$

式(7)用于测定深度视频序列每个位置的强度值, 其中 $g(u, v; \sigma_l^2)$ 为空间域上的二维高斯核函数,

$$g(u, v; \sigma_l^2) = \frac{1}{2\pi\sigma_l^2} \exp\left(-\frac{(u^2 + v^2)}{(2\sigma_l^2)}\right) \quad (8)$$

h_{ev} 和 h_{od} 是时间域中一对正交的一维 Gabor 滤波器。 $h_{ev}(t|\tau, \omega) = \cos(2\pi\omega t)e^{-t^2/2\tau^2}$, $h_{od}(t|\tau, \omega) = \sin(2\pi\omega t)e^{-t^2/2\tau^2}$, $\omega = \tau/4$ 。上述式中 σ_l^2 和 τ 分别控制着函数在空间域和时间域上的检测尺度。与 Harris STIPs 检测一样, 选择响应函数的局部最大值作为深度视频序列的 STIPs。

2.3 STIPs 特征描述

在得到相应 STIPs 之后, 由于 STIPs 在时空域中是孤立的像素点, 深度信息的 STIPs 属性只有其时空坐标和深度值。还需要提取像素点邻域的一些特征信息来对运动做进一步的描述。基于自相似性来编码三维立方体的时空形状, 定义了深度立方体相似特征(DCSF)^[5]。如图 2 所示, 对于检测到的每一个 STIP, 围绕兴趣点定义一个立方体, 把它划分成 $n_{xy} \times n_{xy} \times n_t$ 小立方体。计算包含在每个块的深度像素点的直方图, 标准化之后使得每个直方图的总值为 1, 分别计算块 p 和 q 的直方图, 得到的结果分别为 h_p 和 h_q 。使用巴氏距离来定义相似度:

$$S(p, q) = \sum_{n=1}^M \sqrt{h_p^{(n)} h_q^{(n)}} \quad (9)$$

这里描述的是 2 个块之间的深度关系。 M 表示直方图区间的数目。通过连接所有分块的相似度生成特征向量。不同的空间大小从 1×1 到 $n_{xy} \times n_{xy}$, 有 $n_{xy}(n_{xy}-1)(2n_{xy}-1)/6$ 种可能性。不同的时间大小 1 到 n_t 有 $n_t(n_t+1)/2$ 种可能性。总体来说, 块的个数 N_b 是由立方体所包含的像素决定的, 它的总量接近于 $n_t^2 n_{xy}^3 / 6$ 。为了减少计算的成本, 使用积分直方图来快速计算深度直方图^[6]。将深度像素量化成 M 份, $M = (d_{\max} - d_{\min}) / \Delta d$ 。这里 Δd 是根据识别动作的空间层次而选择的, 比如: $\Delta d = 0.2$ m。

然后, 生成 M 份量化的视频量 Q^n , $n=1, 2, \dots, M$, 对应于 M 份:

$$Q^n(x, y, t) = \begin{cases} 1, & \text{if } (n-1)\Delta d + 1 \leq D(x, y, t) \leq n\Delta d \\ 0, & \text{else} \end{cases} \quad (10)$$

对于每个量化的视频卷 Q^n , 计算一个合成的视频卷 I^n , $n=1, 2, \dots, M$:

$$\begin{cases} r^{(n)}(x, y, t) = r^{(n)}(x, y-1, t) + Q^n(x, y, t) \\ c^{(n)}(x, y, t) = c^{(n)}(x-1, y, t) + r^{(n)}(x, y, t) \\ I^n(x, y, t) = I^n(x, y, t-1) + c^{(n)}(x, y, t) \end{cases} \quad (11)$$

式中: $r^{(n)}(x, y, t)$ 表示 $Q^n(x, y, t)$ 行像素的和; $c^{(n)}(x, y, t)$ 表示 $r^{(n)}(x, y, t)$ 列像素的和; $I^n(x, y, t)$ 表示 $c^{(n)}(x, y, t)$ 时间维度的和。那么只用 7 个加法操作就可以得到一个块第 n 份的直方图的计算:

$$B^{(n)} = \{I^{(n)}(p_8) - I^{(n)}(p_7) - I^{(n)}(p_6) + I^{(n)}(p_5)\} - \{I^{(n)}(p_4) - I^{(n)}(p_3) - I^{(n)}(p_2) + I^{(n)}(p_1)\} \quad (12)$$

在图 2 中已经给出了 $P_1, P_2, \dots, P_8$ 位置的标注。每个视频只计算一次积分视频卷, 并且每块的直方图只需有 $7M$ 次加法操作就可以得到。

3 动作分类器

对所有视频序列的时空兴趣点进行类别统计后, 选用支持向量机(SVM)来完成对测试数据集进行识别。对于本课题研究的深度视频序列来说, 特征数据样本的维数非常高, 为此, SVM 通过映射的方法将特征数据样本点投影到高维空间, 使得在高维空间里找到一个决策曲面的超平面完成线性划分。本实验中采用非线性 SVM 进行数据分类, 式(13)为本实验中所用到的函数:

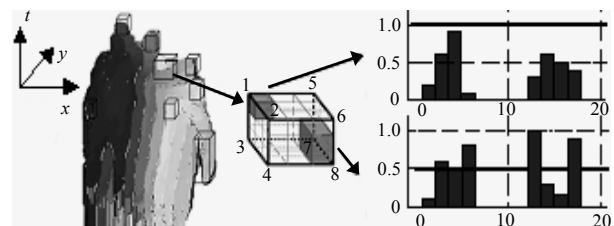


Fig.2 Illustration of extracting DCSF from depth video
图 2 从深度视频提取 DCSF 的图解

$$K(x_i, x_j) = \sum_{i=1}^n \min(x_i, x_j), x_i \geq 0, x_j \geq 0 \quad (13)$$

4 实验及结果分析

在运动识别中选取的数据库不同,往往会影响到识别算法的结果,使用不同的数据库会得到不同的识别结果。但是,相应的实验结果是检测算法适应性和有效性的重要途径。在本实验中,运用 Matlab 编程平台完成兴趣点提取和特征提取,并且使用了台湾林智仁教授开发的 libSVM 分类器^[7]动作的分类。此外,选取了公开的人体运动深度信息数据库 MSR Daily3D^[8]和 MSR Action3D^[9]数据库。这 2 种数据库均由微软研究小组采集,在这里需要强调的是 MSR Action3D 数据库是经过背景去除预处理的,而 MSR Daily3D 运动数据库只是去除了距离较远的部分,因此 MSR Daily3D 运动数据库的识别更具挑战性。

4.1 MSR Daily3D 运动数据库

MSR Daily3D 数据库共包含 16 种动作: drink, eat, read book, call cellphone, write on a paper, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lay down, walk, play guitar, stand up, sit down。每种运动由 10 个人完成,每人做 2 遍,主要涵盖了起居室可能发生的日常行为。

4.1.1 Harris 时空兴趣点检测结果

实验中 Harris STIPs 检测的参数 σ_t^2 和 τ_l^2 分别取 6 和 15, $N_0=400$, 并选取立方体数目为 $n_{xy}=4, n_l=3$, 将所有兴趣点全都映射到一幅深度图像中,其中“sit down”动作的检测结果如图 3 所示。所有动作在 MSR Daily3D 数据库上的识别结果如图 4 所示。

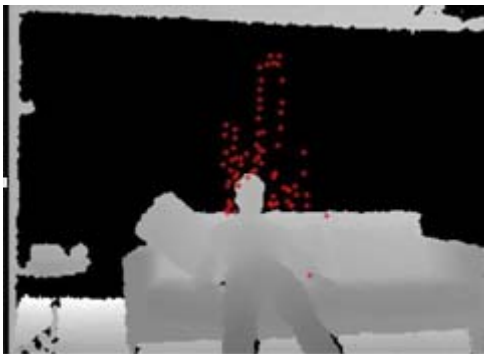


Fig.3 STIPs detected by Harris detector of 'sit down' in MSR Daily3D dataset

图 3 MSR Daily3D 数据库中“sit down”的 Harris 时空兴趣点检测结果

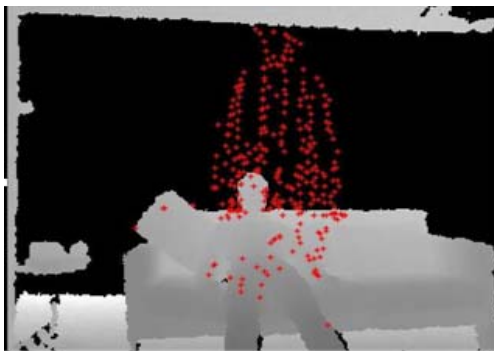


Fig.5 STIPs detected by Gabor filter detector of 'sit down' in MSR Daily3D dataset

图 5 MSR Daily3D 数据库中“sit down”的 Gabor 滤波器时空兴趣点检测结果

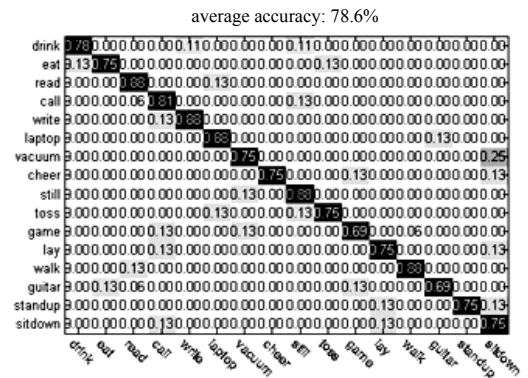


Fig.4 Recognition result on MSR Daily3D dataset with STIPs detected by Harris detector

图 4 Harris 时空兴趣点在 MSR Daily3D 数据库上的识别结果

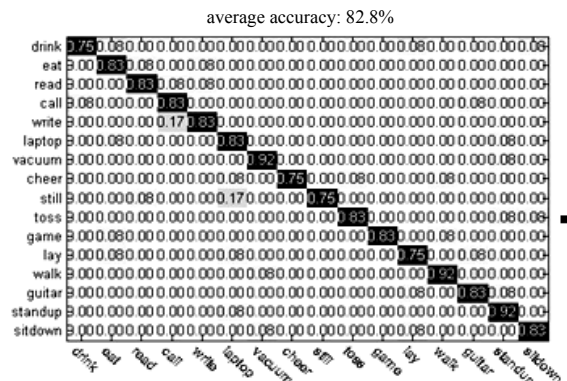


Fig.6 Recognition result on MSR Daily3D dataset with STIPs detected by Gabor filter detector

图 6 Gabor 滤波器时空兴趣点在 MSR Daily3D 数据库上的识别结果

4.1.2 基于 Gabor 滤波器的时空兴趣点检测

取 Gabor 滤波器的空间尺度 $\sigma_l^2=9$, 时间尺度 $\tau=12$, $N_b=400$, 并选取立方体数目为 $n_{xy}=4, n_t=3$, 将整个“sit down”动作序列的 STIPs 全都映射到一幅深度图像中, 如图 5 所示。所有动作在 MSR Daily3D 数据库上的识别结果如图 6 所示。

4.1.3 2 种 STIPs 检测结果分析

在 MSR Daily3D 数据库中, Harris STIPs 方法取得的识别率为 78.6%, 而 Gabor 滤波器 STIPs 方法取得的识别率为 82.8%。因此, 在 MSR Daily3D 数据库中, 后者方法取得较好的识别效果。

4.2 MSR Action3D 运动数据库

MSR Action3D 数据库包含 20 种运动: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, side boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing, pick up & throw。每种运动由 10 个人完成, 每人做 3 遍。

4.2.1 Harris 时空兴趣点检测结果

实验中, $\sigma_l^2=7, \tau_l^2=6$ 进行 Harris STIPs 检测, $N_b=500$ 并选取立方体数目为 $n_{xy}=4, n_t=2$ 。如图 7 所示, “high throw”动作序列兴趣点都映射到一幅深度图像里。图 8 显示了数据库中所有动作的识别结果。

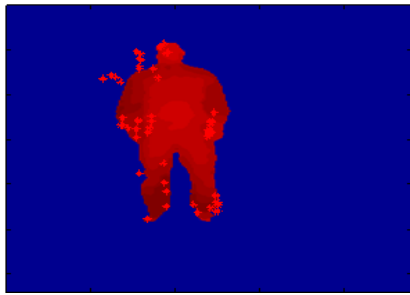


Fig.7 STIPs detected by Harris detector of 'high throw' in MSR Action3D dataset

图 7 MSR Action3D 数据库中“high throw”的 Harris 时空兴趣点检测结果

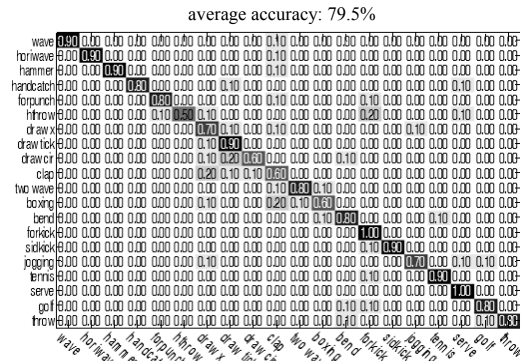


Fig.8 Recognition result on MSR Action3D dataset with STIPs detected by Harris detector

图 8 Harris 时空兴趣点在 MSR Action3D 数据库上的识别结果

4.2.2 基于 Gabor 滤波器的 STIPs 检测

取 Gabor 滤波器的空间尺度 $\sigma_l^2=5$, 时间尺度 $\tau=7$, $N_b=500$, 并选取立方体数目为 $n_{xy}=4, n_t=2$ 。将“high throw”动作序列的时空兴趣点全都映射到一幅深度图像中, 如图 9 所示。所有动作在 MSR Daily3D 数据库上的识别结果如图 10 所示。

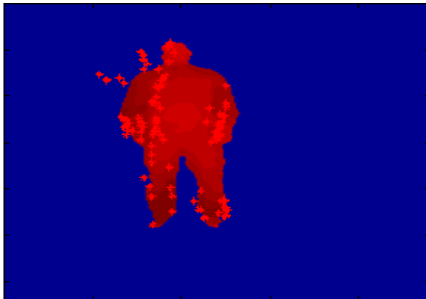


Fig.9 STIPs detected by Gabor filter detector of 'high throw' in MSR Action3D dataset

图 9 MSR Action3D 数据库中“high throw”的 Gabor 滤波器时空兴趣点检测结果

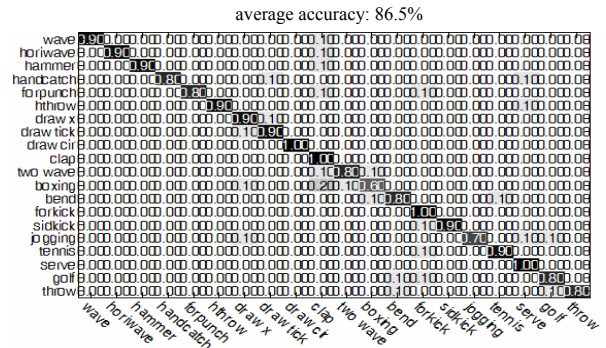


Fig.10 Recognition result on MSR Action3D dataset with STIPs detected by Gabor filter detector

图 10 Gabor 滤波器时空兴趣点在 MSR Action3D 数据库上的识别结果

4.2.3 2 种 STIPs 检测结果分析

在 MSR Action3D 数据库中, Harris STIPs 方法取得的识别率为 79.5%, 而 Gabor 滤波器 STIPs 方法取得的识别率为 86.5%。因此, 在 MSR Action3D 数据库中, 后者同样取得较好的识别效果。最后, 将 Harris STIPs 检测方法和 Gabor 滤波器 STIPs 检测方法作对比, 如表 1 所示。利用 Gabor 滤波器 STIPs 方法在深度视频数据库中的

识别率相比 Harris STIPs 检测方法要高出许多。

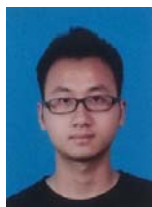
5 结论

本文在深度信息的基础上研究了有关人体运动识别的方法,展示了从深度视频中提取 STIPs 的算法和围绕着兴趣点的局部 3D DCSF 的计算。将人体运动识别的 2 个主要检测方法在深度视频库中进行了比较。同样,本文研究方法也存在很大的拓展性,当骨骼关节等^[10]精确信息可以利用时,DCSF 结合关节特征会带来更加精确的识别效果。

参考文献:

- [1] 雷庆,李绍滋. 动作识别中局部时空特征的运动表示方法研究[J]. 计算机工程与应用, 2010,46(34):7-10. (LEI Qing, LI Shaozi. Research on local spatio-temporal features for action recognition[J]. Computer Engineering and Applications, 2010,46(34):7-10.)
- [2] 辛彦哲,冯辉,杨涛,等. 基于特征子空间的自适应多视角目标跟踪算法[J]. 太赫兹科学与电子信息学报, 2012,10(3):319-324. (XIN Yanzhe, FENG Hui, YANG Tao, et al. Adaptive subspace tracking algorithm on multi-view videos[J]. Journal of Terahertz Science and Electronic Information Technology, 2012,10(3):319-324.)
- [3] LAPTEV I. On space-time interest points[J]. International Journal of Computer Vision, 2011,64(2-3):432-439.
- [4] DOLLAR P, RABAU D V, COTTRELL G, et al. Behavior recognition via sparse spatio-temporal features[C]// 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China:[s.n.], 2005:65-72.
- [5] XIA L, AGGARWAL J K. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2013,9(4):2834-2841.
- [6] PORIKLI F. Integral histogram: a fast way to extract histograms in Cartesian spaces[C]// 2005 IEEE Computer Society Conference on Computer Vision & Pattern Recognition. San Diego, USA:[s.n.], 2005:829-836.
- [7] CHANG C, LIN C. LIBSVM: A library for support vector machines[J]. Acm Transactions on Intelligent Systems & Technology, 2001,12(3):389-396.
- [8] WANG J, LIU Z, WU Y, et al. Mining actionlet ensemble for action recognition with depth cameras[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, Rhode Island:[s.n.], 2012:1290-1297.
- [9] LI W, ZHANG Z, LIU Z. Action recognition based on a bag of 3D points[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. San Francisco, USA:[s.n.], 2010: 9-14.
- [10] YANG X, TIAN Y. Eigen joints-based action recognition using Naïve-Bayes-Nearest-Neighbor[J]. Perceptual & Motor Skills, 2012,38(3c):14-19.

作者简介:



陈 力(1993-), 男, 安徽省六安市人, 在读硕士研究生, 主要研究方向为信号与信息处理. email: novadechen0813@163.com.

王俊凯(1992-), 男, 四川省攀枝花市人, 在读硕士研究生, 主要研究方向为实时信号处理、FPGA 高速信号处理、嵌入式系统。

张玉玺(1985-), 男, 河南省驻马店市人, 博士, 讲师, 主要研究方向为雷达信号处理等。

表1 文中2种识别方法对比

dataset	recognition rate	
	Harris detection	Gabor filter detection
MSR Daily3D dataset	78.6%	82.8%
MSR Action3D dataset	79.5%	86.5%