

基于时空兴趣点的人体行为识别

付朝霞^{1,2}, 王黎明¹

(1 中北大学电子测试技术重点实验室 & 仪器科学与动态测试教育部重点实验室, 信息通信工程学院, 山西 太原 030051;
2 山西省委党校 信息网络教研部, 山西 太原 030006)

摘 要: 提出一种基于时空兴趣点的人体行为识别算法, 该算法从行为视频中提取丰富的时空兴趣点, 然后在数据空间采用基于高斯混合模型的聚类方法完成对特征点的分类. 在动作识别过程中, 采用平均 Hausdorff 距离来衡量序列间匹配的相似性, 提高运算效率. KTH 数据库上的实验证明了该算法的有效性和鲁棒性.

关键词: 行为识别; 特征提取; 时空兴趣点; 隐马尔可夫模型; 平均 Hausdorff 距离

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-7180(2013)08-0028-03

Human Action Recognition Based on Space-time Interest Point

FU Zhao-xia^{1,2}, WANG Li-ming¹

(1 Science and Technology on Electronic Test & Measurement Laboratory and Key Laboratory of Instrumentation Science & Dynamic Measurement(Ministry of Education), Information and Communication Engineering Institute, North University of China, Taiyuan 030051, China;
2 Party School of Shanxi Provincial Committee of the C. P. C, Taiyuan 030006, China)

Abstract: Human action analysis and recognition are increasingly attracting much attention from computer vision and pattern recognition researchers. This paper presents a human action recognition based on space-time interest point algorithm. The algorithm detects dense space-time interest points from action videos and classifies these feature points using cluster method based on Gaussian mixture model in the data space. In the process of action recognition, a matching-based approach with the mean Hausdorff distance is measured the similarity between image sequences to improve the operation efficiency. The experiments on the KTH database prove the effectiveness and robustness of the algorithm.

Key words: action recognition; feature extraction; space-time interest point; Hidden Markov models; mean Hausdorff distance

1 引言

人体运动的视觉分析作为计算机视觉中最活跃的研究领域之一, 近年来已引起计算机科研人员极大的兴趣, 在许多工业民用及军事领域有着广泛的应用前景^[1-2].

在一个人体运动的视觉分析系统中大体可分为四个不同的功能单元: 人体目标检测、人体目标跟踪、人体目标分类、人体行为识别和理解. 其中人体

行为识别是在成功实现图像序列中的人体检测, 完成特征提取和跟踪的基础上进行的, 属于更高层次的视觉任务, 是当前计算机视觉研究领域备受关注并最具挑战的研究方向之一^[3]. 但由于目前的计算机视觉方法仍然存在尚不尽人意的地方, 而寻找鲁棒的行为图像特征表征方法是行为识别成功的关键. 同时, 人体运动序列中姿态随时间不断的变化, 也给人体行为准确的分类和识别带来很大的困难. 当前比较流行的 HMM^[4]方法虽然有较强的时间建

模能力,但其计算的复杂度较大影响了分类的效率。

本文提出一种基于时空兴趣点的人体行为识别算法,采用 3D Harris 角点的时空兴趣点提取不同行为的时空特征,然后采用高斯混合模型对提取的运动特征进行分类,训练识别过程则采用平均 Hausdorff 距离^[5]的几何特征方法完成相似性配准。

2 运动识别系统

本文的运动识别系统由提取时空特征模块、特征分类模块和运动识别模块组成。

2.1 时空特征的提取与表示

由于全局特征(如运动轨迹、光流、形状、时空体等)仅在背景简单、摄像机固定、图像分辨率较高等特定的应用场景中识别效果比较理想,但在部分遮挡和背景噪声干扰等因素影响下效果欠佳。因此为了弥补全局特征的不足,基于局部运动的时空兴趣点特征方法应运而生。它反映了视频中图像变化显著的区域,具有很好的特征描述和类别区分能力。最常用的时空兴趣点是 Laptev 等^[6]将 2D Harris 角点^[7]扩展到时空领域,即提出一种局部 3D Harris 角点检测算子应用于行为识别,本文也采用此兴趣点,下面简要介绍其计算过程。

在时空序列图像中,我们可以构建平滑图像序列 $I(x, y, t)$ 的线性尺度空间表征:

$$L(x, y, t, \sigma^2, \tau^2) = g((x, y, t, \sigma^2, \tau^2) * I(x, y, t)) \quad (1)$$

其中 g 为时空可分离的高斯核,定义为

$$g(x, y, t, \sigma^2, \tau^2) = \frac{\exp(-(x^2 + y^2)/2\sigma^2 - t^2/2\tau^2)}{\sqrt{(2\pi)^3 \sigma^4 \tau^2}} \quad (2)$$

这里 σ^2 和 τ^2 为独立的空间和时间均方差。

与 2D Harris 角点类似,使用高斯函数与平滑图像序列的时空二阶矩的卷积 μ 的响应函数寻找 3D Harris 角点。卷积 μ 定义为

$$\mu = g(x, y, t, \sigma^2, \tau^2) * \begin{bmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{bmatrix} \quad (3)$$

其中, $L_x = \partial_x(g * I)$, $L_y = \partial_y(g * I)$, $L_t = \partial_t(g * I)$ 分别为平滑图像序列的一阶导数。

在检测兴趣点时,扩展的 Harris 角点的响应函数为

$$R = \det(\mu) - k * \text{trace}^3(\mu) \quad (4)$$

式中, k 通常取值为 0.005。

2.2 特征分类

通过有效的对时空兴趣点进行分类,可以为单人行行为识别模型提供了可靠的观测特征。高斯混合模型是一种半参数的密度估计方法,能够很好地刻画参数空间中数据的空间分布及其特性,因此我们采用流行的高斯混合模型对运动特征进行分类。

假定来自运动某类的一组特征向量 $X = \{x_1, x_2, \dots, x_N\}$ 为 N 个观察到的数据向量组成的集合,单个采样 x_i 的高斯混合分布的密度函数为

$$p(x_i; \Theta) = \sum_{j=1}^k \omega_j * N(x_i; u_j, \Sigma_j) \quad (5)$$

这里 $\Theta = \{\omega_j, u_j, \Sigma_j\}_{j=1}^k$, Σ_j 是协方差矩阵, k 是高斯分量的数目。 ω_j 是第 j 个高斯分量的权,且 $\sum_{j=1}^k \omega_j = 1, 0 \leq \omega_j \leq 1$ 。

当所要估计的参数是非随机的未知常量时(或者待估计参数是随机的,但先验密度未知时),则通过使密度函数最大化来确定参数的值,即根据 Θ 的极大似然估计进行分类:

$$c = p(X; \Theta) = \arg\max_{\Theta} \sum_{i=1}^N \log p(x_i; \Theta) \quad (6)$$

K-means 算法是基于 GMM 模型的 EM 求解算法的一种特例,所以我们可以利用 K-means 聚类算法检测时空点区,每个聚类中心定义为 $m_i = \frac{1}{n_i} \sum_{k=1}^{n_i} j_k$, 即把 n_i 个点分为 c_i 个类,这样就把每个检测到的兴趣点被划分到唯一的类成员。

2.3 运动识别

假设一个分类为 c_i 的特征模型序列数据集 $X = \{x_1, x_2, \dots, x_N\}$, 其数据特征为 $D = \{f^d = (x^d, y^d, t^d, \sigma^d, \tau^d, c^d) | t^d \in (t_1, t_2)\}$, $t^d \in (t_1, t_2)$ 表示检测的时间段, $Y = \{y_1, y_2, \dots, y_N\}$ 为给定的一个时间连续的测试序列。模型与数据之间的匹配定义为找到满足最小值问题的加权总和的距离,使模型特征 f^m 和数据特征 f^d 匹配。

基于平均 Hausdorff 距离的配准方法本身具有实现简单、速度快的特点。由于基于 GMM 模型的方法反映了数据的空间分布和特性,忽略了帧间的时间关系。为了处理不同序列中的时空变换,我们使用平均 Hausdorff 距离来衡量匹配的相似性。

平均 Hausdorff 距离定义为

$$S(f^m, f^d) = \text{mean}(\min(|\tilde{M}_1 - \tilde{M}_2|)) \quad (7)$$

其中 $\tilde{M}_1 = \{f^m | t^m \in (t_1, t_2)\}$,

$\tilde{M}_2 = \{f^d | t^d \in (t_1, t_2)\}$ 。

这样,序列 X 与序列 Y 相似性距离为

$$d_h(f^m, f^d) = \frac{1}{2} [S(f^m, f^d) + S(f^d, f^m)] \quad (8)$$

因此,对测试运动序列分类也就是最小化相似性距离 $d_h(f^m, f^d)$,即 $\text{argmind}_h(f^m, f^d)$,可以使用标准的高斯-牛顿优化方法对这个结果进行优化.

3 实验分析

本文利用 KTH^[8] 数据库进行实验评估. KTH 是相对比较复杂的 2D 单人行为库, 包含有 6 种常见的行为类别: 行走(walking)、慢跑(jogging)、跑步(running)、拳击(boxing)、击掌(hand clapping)、挥手(hand waving). 每种行为由 25 个个体在 4 个不同场景中分别完成. 这 4 个场景包括: 室外(s1)、室外尺度变化(s2)、室外个体穿着变化(s3)、室内(s4). 当前该数据库共包含 2 391 段视频序列, 视频的帧速为每秒 25 帧, 帧的大小为 160×120 . 实验中, 对于每个场景, 我们在数据库中随机选取 16 个人的行为作为训练集, 剩下的 9 个人的行为作为测试集.

图 1 是从行走的人中检测出时空兴趣点. 图 2 是采用 K-means 算法对时空兴趣点分类的结果, 聚类中心个数为 50, 其中不同颜色代表不同的类别. 表 1 所示为 KTH 库中行为序列识别的混淆矩阵. 从表 1 中, 可以看出对于“慢跑”、“跑步”区分度不大的



图 1 时空兴趣点的检测结果



图 2 K-means 算法对时空兴趣点分类的结果

动作, 相对辨识度稍低, 而其他较易区分的动作, 取得了令人满意的识别结果.

表 2 中列出了本文算法与文献[8]算法在相同训练集和测试集上的识别结果. 由表 2 可见, 该算法的系统的识别率高于文献[8]中的算法.

表 1 混淆矩阵

姿势	walking	jogging	running	boxing	waving	clapping
walking	1	0	0	0	0	0
jogging	0.04	0.92	0.04	0	0	0
running	0	0.04	0.96	0	0	0
boxing	0	0	0	1	0	0
waving	0	0	0	0.01	0.98	0.01
clapping	0	0	0	0	0	1

表 2 算法识别率比较

Method	Laptev, et al ^[6]	ours
Classification Accuracy	91.8%	97.7%

4 结束语

本文提出的基于时空兴趣点的人体行为识别算法在 KTH 数据库中取得了较好的识别性能, 时空兴趣点的特征方法弥补全局特征的不足, 具有很好的特征描述和类别区分能力, 而使用高斯混合模型进行分类的方法和平均 Hausdorff 距离衡量匹配的相似性, 进一步提高了识别的准确率和运算效率. 本文算法并不局限于单人行为分析, 可以扩展到多人交互行为识别. 在多人交互行为过程中人与人之间存在严重的遮挡和自遮挡, 在多视角场景下合理利用深度信息, 可以有效提高特征提取的准确性, 这也是我们未来的研究重点.

参考文献:

- [1] Gavrilu D. The visual analysis of human movement: a survey[J]. Computer Vision and Image Understanding, 1999, 73(1): 82-98.
- [2] Xu F, Liu X, Fujimura K. Pedestrian detection and tracking with night vision [J]. IEEE Transactions on Intelligent Transportation Systems, 2005, 6(1): 63-81.
- [3] Turaga P, Chellappa R, Subrahmanian V S, et al. Machine recognition of human activities: a survey [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1473-1488.

(下转第 35 页)

造的本文描述子在异源图像上实现匹配,理论上是可行的.图4(c)的实验结果验证了上述的理论分析,在21组的直线匹配对中正确匹配对数为16组,正确匹配率为76.19%.

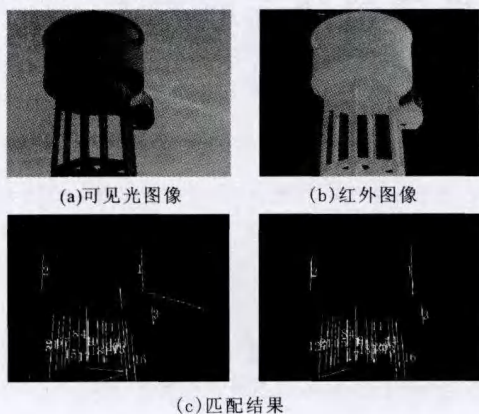


图4 异源图像匹配

6 结束语

本文提出了一种基于 Haar 小波响应的直线匹配算法.实验结果表明本文提出的算法在直线遮挡达到60%时、当图像缩放达4倍时,仍能实现直线匹配;能在旋转、仿射变换条件下实现精确匹配.同时,本文提出的算法在红外与可见光图像中也能实现有效匹配.

参考文献:

[1] Watanabe T, Tanaka T. Vein authentication using color information and image matching with high performance on natural light[C]//Proceedings of International Joint

Conference 2009. Fukuoka: Fukuoka International Congress Center, 2009: 3625—3629.

- [2] Yang M Y, Cao Y, McDonald J. Fusion of camera images and laser scans for wide baseline 3D scene alignment in urban environments[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011 (66): 52—61.
- [3] Cao Y, McDonald J. Improved feature extraction and matching in urban environments based on 3D viewpoint normalization[J]. Computer Vision and Image Understanding, 2012 (116): 86—101.
- [4] Guan T, Wang C. Registration based on scene recognition and natural features tracking techniques for wide-area augmented reality systems[J]. IEEE Transactions on Multimedia, 2009, 11(8): 1393—1406.
- [5] Bay H, Andreas E. SURF: speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346—359.
- [6] Canny J. A computational approach to edge detect[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679—698.

作者简介:

朱英宏 男,(1983—),博士研究生.研究方向为红外图像处理.

李俊山 男,(1956—),博士,教授.研究方向为图像处理与目标识别、网络信息安全、电子对抗模拟与仿真.

赵怀勋 男,(1960—),硕士,教授.研究方向为图像匹配.

杨威 男,(1982—),博士,讲师.研究方向为红外图像处理.

胡双演 男,(1978—),博士,讲师.研究方向为红外图像处理.

(上接第30页)

[4] Yamato J, Ohya J. Recognizing human action in time-sequential images using hidden markov models[C]//Proceedings of IEEE Computer Society Conference Computer Vision and Pattern Recognition. Champaign, Illinois, 1992: 379—385.

[5] 张见威,韩国强,沃焱.基于边界距离场互信息的图像配准方法[J].通信学报,2006,27(7):87—93.

[6] Laptev I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2): 107—123.

[7] Harris C, Stephens M. A combined corner and edge

detection[C]//Proceedings of the Fourth Alvey Vision Conference. Manchester, British, 1988: 147—151.

[8] KTH database [EB/OL] (2006—11—06). <http://www.nada.kth.se/cvap/actions>.

作者简介:

付朝霞 女,(1978—),博士研究生.研究方向为计算机视觉与模式识别.

王黎明 男,(1974—),教授.研究方向为图像分析与理解、模式识别与机器学习.