文章编号：2096-1618(2019)04-0358-07

融合时空兴趣点和多元广义高斯混合模型的人体动作识别

何冰倩， 魏 维， 宋岩贝， 高联欣， 张 斌
(成都信息工程大学计算机学院，四川 成都 610225)

摘要：人体动作识别近年作为计算机视觉领域的热点研究方向，被广泛用于人机交互、虚拟现实等领域。针对传统人体动作识别算法中提取特征时冗余点过多、忽略图像数据的关联性等问题，提出一种融合时空兴趣点和结合定点估计的多元广义高斯混合模型(MGGMMs)的人体动作识别方法，通过过滤冗余特征点和利用多元广义高斯混合模型实现了特征点的有效提取以及对数据关联性的充分利用。以改进的 Harris-Laplace 算法和 3D-SIFT 描述子提取视频序列的特征点，利用 BOW 模型进行视觉词聚类，最后通过改进的多元广义高斯混合模型进行建模和分类。在 KTH 公开数据集上进行实验，实验结果表明提出的人体动作识别方法能够对视频中人体动作进行有效识别和分类。

关 键 词：动作识别；时空兴趣点；Harris-Laplace；MGGMMs；特征提取
中图分类号：TP391.41 **文献标志码**：A
doi：10.16836/j.cnki.jcuit.2019.04.006

0 引言

人体动作识别要解决的主要问题是如何将摄像机或传感器采集到的视频序列通过分析和处理[1]，使计算机能够“理解”视频中人类的动作和行为[2]，对安全监控、娱乐方式等方面都有重要的研究意义，而基于视频的人体动作识别在人机交互[3]、虚拟现实、智能家居设备等领域也都有广泛应用。人体动作识别正逐渐成为计算机视觉领域最活跃的研究热点之一[4-5]，然而基于视频的人体动作识别仍然存在巨大挑战性，其原因主要是两个方面，一方面是视频环境因素，另一方面是动作类别本身的复杂度。比如视频光照的变化、摄像机的抖动、视角的变化等都是属于视频环境因素。而对于动作类别本身，主要是类间和类内差异问题。比如“慢跑”、“散步”和“跑步”这3个不同类别，由于动作速度等原因，造成不同类别间差异较小的问题；而对于相同动作，由于视角等原因也会造成相同类别的动作有较大的差异问题。

近年来，许多国内外学者对人体动作识别任务进行了大量的研究和实验，大致可以将人体动作识别分为3个步骤：特征提取、特征表示和动作识别。特征提取的方式主要有如下几种[2,6]：提取静态特征基于姿态估计的方法，提取运动信息的动态特征基于底层跟踪的方法，提取光流信息的动态特征基于光流计算的方法，提取时空特征基于图像处理的方法，提取描述性特征基于机器学习的方法。较于形状、运动轨迹等这些全局特征，基于局部时空兴趣点的特征能较好地弥补由于视频遮挡、相机抖动等原因对动作识别造成的不利影响。因此，这类方法在动作识别领域取得了广泛应用[7-10]。然而，比较原始的基于局部时空兴趣点的方法一方面没有充分利用局部特征间的关系，另一方面提取的特征冗余点较多。Dollar 等[11]提出了分别在视频序列的空间维和时间维进行 Gabor 滤波操作，让检测到的时空兴趣点满足有效且稳定的需求。Wang 等[12]利用多尺度时空上下文特征提高时空兴趣点特征的描述能力，以此来利用时空兴趣点的上下文关系，并将该方法应用于动作识别。文献[13]在上述问题和不足的背景下提出了能够提高时空兴趣点检测精度改进的 Harris-Laplace 算法，该方法通过空间尺度选择、时间尺度抑制等方法去除冗余点，从而提高检测精度。基于概率统计的动作识别方法也一直是学者们研究的热点方向[14-16]。Niebles 等[17]提出了一种无监督的动作识别方法，该方法利用概率潜在语义分析(pLSA)模型和隐含狄利克雷分布(LDA)来实现自动学习时空兴趣点的概率分布，并与人体动作类别建立相对应的关系，从而实现人体动作识别分类。文献[18]利用广义高斯混合(GGM)和回归技术建立了人体动作识别分类的计算模型，广义高斯混合模型(GGMMs)由于其比高斯分布更具有数据表征的灵活性，使该混合模型较之前的算法在识别准确率上有一

收稿日期：2018-12-19
基金项目：四川省教育厅重点科研项目(17ZA0064)

定增长,但是在多元分析的情况下,上述方法仍然忽略了人体动作数据的关联性。

因此,针对上述方法中存在的问题和不足,提出了一种融合时空兴趣点和多元广义高斯混合模型的人体动作识别方法。首先利用 Harris-Laplace 算法提取视频序列中的时空兴趣点,再利用 3D-SIFT 描述子对初步提取的时空兴趣点进行处理,得到向量化的描述子。再利用 BOW(bag-of-words)算法和改进后多元广义高斯混合模型对人体动作进行建模和分类。在公开数据集 KTH 上进行人体动作识别实验,实验结果表明,提出的融合时空兴趣点和多元广义高斯混合模型的人体动作识别方法能够有效识别视频序列中不同类别的人体动作。

1　动作识别模型

提出的融合时空兴趣点和多元广义高斯混合模型的人体动作识别方法主要包含 4 个部分:(1) 时空兴趣点的提取和过滤;(2) 利用 3D-SIFT 提取特征描述子;(3) 通过 BOW 算法构建视觉单词,并利用 K-Means 算法进行初步聚类;(4) 利用改进后的多元广义高斯混合模型(MGGMMs)对整体人体动作进行建模分类。

1.1　时空兴趣点检测

传统 Harris 兴趣点检测算法具有旋转不变性,但是不包含尺度信息,从而检测出来的兴趣点多且冗杂。Harris-Laplace 算法结合了传统的 Harris 算法和 Laplace 尺度空间,从而实现了该算法的尺度不变性,但是在检测任务中,仍然会出现冗杂无用的兴趣点。因此,利用改进后的 Harris-Laplace 算法对数据集中的视频序列进行时空兴趣点检测,并对检测到的时空兴趣点进行去除冗余点操作,这使获取到的时空兴趣点尽可能有效准确,从而为后期的特征表示和动作分类提供更为准确有效的数据支持。图 1 是传统 Harris-Laplace 兴趣点检测和本文的检测算法的优化对比图示。示例来自数据集 KTH 的同一测试者的 6 个不同动作(“boxing”、“hand clapping”、“hand waving”、“jogging”、“running”、“walking”)。由图 1 可以看出,算法检测出的时空兴趣点主要集中在人体的运动部位,且一定程度上去除了冗杂无用的兴趣点。

时空兴趣点检测的具体算法如下:

首先对输入的视频序列进行多尺度兴趣点检测。将视频序列与变换高斯核函数滤波器在不同尺度上进行卷积操作,得到尺度空间 $S(x,y,\sigma)$,计算公式为

(a)原算法的时空兴趣点检测

(b)文中算法的时空兴趣点检测

图 1　KTH 数据集多尺度时空兴趣点检测优化对比

$$S(x,y,\sigma)=G(x,y,\sigma)I\otimes(x,y) \tag{1}$$

其中,$I(x,y)$是输入视频序列图像,σ 为高斯核函数的尺度因子。多尺度高斯核函数 $G(x,y,\sigma)$表示为

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}} \tag{2}$$

然后,利用公式(3)计算每个尺度空间图像上各像素点 (x,y) 的多尺度响应值 R,并设置阈值 T。此处阈值 T 的主要作用是控制提取到的角点个数,若阈值过大,则会造成冗余点较多形成噪声的问题,而影响识别结果;若阈值过小,则不足以提取动作的特征,造成无法区分较为相似动作的问题,从而减低识别准确率。即响应值 R 越大,则说明对应该像素点越有可能是有效特征。

$$R=\det(M)-m\times\operatorname{trace}^2(M)>T \tag{3}$$

$$M=\sigma_D{}^2g(\sigma_I)\otimes\begin{pmatrix}L_x^2(x,\sigma_D) & L_xL_y(x,\sigma_D)\\ L_xL_y(x,\sigma_D) & L_y^2(y,\sigma_D)\end{pmatrix} \tag{4}$$

其中,m 表示一个常数,取值范围一般为0.04 ~ 0.06,实验中取值为0.04。$\boldsymbol{M}$ 为多尺度自相关矩阵,σ_D、σ_I 分别表示微分和积分尺度,它们一般有 $\sigma_I=\alpha\sigma_D$ 这样的关系,α 在实验中取值0.7。L 是尺度空间函数。

然后是构建选择矩阵。先构建候选兴趣点响应矩阵 $\boldsymbol{C}(x,y)$,以及候选兴趣点尺度矩阵 $\boldsymbol{K}(x,y)$,并将 $\boldsymbol{C}$、$\boldsymbol{K}$ 2 个矩阵初始化为 0。然后再将检测到的时空兴趣点的响应值赋值给 $\boldsymbol{C}(x,y)$,和对应的尺度赋值给 $\boldsymbol{K}(x,y)$。

最后是筛选时空兴趣点。先对候选时空兴趣点矩阵进行像素统计滤波操作,像素统计滤波器的半径

为 n,在实验中设置为 $n=3$。然后在 3×3 的比较窗口内比较中心点和邻域内候选兴趣点响应值 R,得到窗口范围内第二大响应值SecMax_k,并以此来对第一步中从视频序列帧中每个像素点 (x,y) 获得的不同尺度下的兴趣点 (x_i,y_i) 进行筛选,即当满足如下约束条件(5)时,保留该兴趣点,若不满足,则将该点对应在矩阵中的值置为 0。

$$\begin{cases}\mathrm{SecMax}_k > T \\ (x_i,y_i) \in C(x,y)\end{cases} \tag{5}$$

其中,T 为控制兴趣点数目的阈值,在实验中,阈值 T 取值 17。最后得到筛选后的候选时空兴趣点矩阵。

通过上述步骤,最终得到的时空兴趣点矩阵较之前直接提取的兴趣点矩阵减少了冗余的兴趣点,提高了兴趣点的检测精度。

1.2 提取特征描述子

在得到视频序列的特征点后,需要进一步对其进行描述子提取,即将特征点生成对应的特征向量。基于图像梯度的描述子已经被应用于一些识别任务中[17],尽管这类描述子运算量小,但是既不提供时域上的尺度不变性也不提供空域上的尺度不变性,无法满足存在有相对运动的视频序列的复杂要求。3D-SIFT较于 2D-SIFT 描述子,增加了第三个维度-时间 t,在对尺度空间下特征点进行梯度运算时,除了需要计算特征点的水平方向和垂直方向的梯度,还需要计算时间 t 方向上的梯度,更为符合人体动作识别任务,因为人体动作在时间序列上具有关联性,因此选用 3D-SIFT描述子来生成特征向量。

首先对检测到的特征点邻域的高斯图像进行梯度计算,获得水平方向、垂直方向、时间 t 方向上的梯度幅值和梯度方向,窗口大小选择为 16×16×16。然后以候选时空兴趣点为特征点中心,通过二维高斯权重函数在视频序列中做三个方向的梯度集成,最后得到 4×4×4×8×8 维的特征向量。

1.3 构建视觉词袋

BOW 视觉词袋模型在文本分类、图像分类等分类任务领域有广泛的应用[19-20]。为了便于后续对不同类别的人体动作进行正确的分类,需要利用 BOW 模型对视频序列提取到的描述子进一步进行特征表示。

在 BOW 模型中,利用 K-Means 聚类算法对上一步提取到的特征描述子进行聚类,聚类中心称作视觉单词(visual words),这一步骤将特征向量转化为视觉单词,从而得到概率直方图的表示。由于后续将利用改进的多元广义高斯混合模型对特征向量进行建模和分类,所以直接使用传统的 K-Means 聚类算法进行聚类。

首先利用 K-Means 聚类算法计算所有的特征描述子与 k 个聚类中心的距离,再将其映射到欧氏距离最近的聚类中心,然后构造出特征描述子词典(vocabulary)。利用直方图统计视频序列图像中每一类特征描述子的出现频率,最后得到频率直方图。在实验中,k 的值设置了 3 个标准{200,500,600}。

1.4 多元广义高斯混合模型

基于视频的人体动作不仅包含视频序列的外观信息,还包含更为重要的时间信息(运动信息)。广义高斯混合模型(GGMMs)因为其分布比高斯分布更好的数据灵活性在很多图像和视频处理中被广泛应用[21]。但是在多元分析的情况下,广义高斯混合模型计算简单的代价是忽略了完整协方差矩阵的完整性,同时,由于广义高斯分布的假设是数据的特征之前相互独立,使该模型忽略了视频处理中数据的关联性。因此,为了充分利用视频序列的时空信息,采用改进后的多元广义高斯混合模型对人体动作进行建模和分类。

D 维的多元广义高斯分布由均值向量 $\mu \in R^D$、协方差矩阵 $\Sigma \in R^{D\times D}$、形状参数 β 组成,其单分布的概率密度函数定义如下:

$$p(X \mid \Sigma;\beta;\mu) = \frac{\Gamma\left(\frac{D}{2}\right)}{\pi^{\frac{D}{2}}\Gamma\left(\frac{D}{2\beta}\right)2^{\frac{D}{2\beta}}} \times \frac{\beta}{\theta^{\frac{d}{2}}\left|\Sigma\right|^{\frac{1}{2}}} \times \exp\left[-\frac{1}{2\theta^{\beta}}(X-\mu)^{T}\Sigma^{-1}(X-\mu)^{\beta}\right] \tag{6}$$

其中 θ 是尺度参数,β 是形状参数,且 $\beta>0$。β 的取值影响高斯分布的形状变化。

多元广义高斯混合模型(MGGMMs)的一般形式如下:

$$f(X \mid \Theta) = \sum_{j=1}^{N} p_j p(X_j \mid \Sigma_j;\beta_j;\mu_j) \tag{7}$$

其中,$\forall j, p_j>0$;$\sum_{j=1}^{N} p_j = 1$,Θ 是所有的参数集合,$\Theta = \{\Sigma_1,\cdots\Sigma_j,\cdots,\Sigma_N;\beta_1,\cdots\beta_j,\cdots,\beta_N;\mu_1,\cdots\mu_j,\cdots,\mu_N;\}$,即对于模型的第 j 个部分,都有一个 Θ_j 的参数集与其对应。$N=\theta\Sigma$。

基于定点估计的方法弥补了广义高斯混合忽略全协方差矩阵的完整性这一缺陷,且对具有关联性的数据友好,因此对于 MGGMMs 模型的参数估计,利用基于定点估计方法的最大似然估计和 EM(expectation-maximization)算法相结合的方式,对数似然函数表示如下:

$$\log(X \mid \Theta) = \sum_{i=1}^{T} \log[f(X_i \mid \Theta)] \tag{8}$$

其中,$\{X_1,\cdots,X_i,\cdots,X_T\}$ 是尺寸为 D 的 T 个特征向量(观测向量)的随机样本。

由于 EM 算法对初始值很敏感,不同的初始值对聚类结果的影响较大,因此利用 BOW 模型的 K-Means 参数来初始化模型参数。改进后的 MGGMMs 模型的具体算法步骤如下:

Step1　模型参数初始化。连接人体动作识别模型的上一步,利用 BOW 模型中 K-Means 算法最后一次迭代的参数来初始化 MGGMMs 模型参数,然后利用矩量法对初始聚类后的各簇进行初始化。

Step2　参数更新。重复该步骤直到对数似然函数 $\log(X \mid \Theta)$ 收敛。

(1) E 步骤。计算概率公式如下:

$$p(j \mid X_i) = \frac{p_j p(X_i \mid \sum_j;\beta_j;\mu_j)}{\sum_{n=1}^{N} p_n p(X_i \mid \sum_n;\beta_n;\mu_n)} \tag{9}$$

(2) M 步骤。

均值估计:

$$\hat{\mu}_j = \frac{\sum_{i=1}^{T} p(j \mid X_i) \mid X_i - \mu_j \mid^{\beta_j - 1} X_i}{\sum_{i=1}^{T} p(j \mid X_i) \mid X_i - \mu_j \mid^{\beta_j - 1}} \tag{10}$$

每个簇的协方差估计,首先根据公式(11)规范化数据集:

$$X_n = X - \mu_j \tag{11}$$

再利用公式(12)、(13)进行协方差矩阵估计:

$$\hat{\sum}_{d+1} = f(\sum_k) \tag{12}$$

$$f(\sum) = \sum_{i=1}^{T} \frac{d}{u_i + u_i^{1-\beta} \sum_{i \neq j} u_j^{\beta}} X_i X_i^T \tag{13}$$

尺度参数:

$$\hat{\theta} = \left[\frac{1}{T}\sum_{i=1}^{T} (u_i)^{\beta}\right]^{\frac{1}{\beta}} \tag{14}$$

其中,

$$u_i = X_i^T \sum^{-1} X_i \tag{15}$$

形状参数估计:

$$\hat{\beta}_{d+1} = \hat{\beta}_d - \frac{\alpha(\hat{\beta}_d)}{\alpha'(\hat{\beta}_d)} \tag{16}$$

其中,

$$\alpha(\beta) = \frac{DT}{2\sum_{i=1}^{T}\mu_i^{\beta}} \sum_{i=1}^{T} [u_i^{\beta} \ln(u_i)] - \frac{DT}{2\beta}\left[\psi\left(\frac{D}{2\beta}\right) + \ln 2\right] - T - \frac{DT}{2\beta}\ln\left(\frac{\beta}{DT}\sum_{i=1}^{T} u_i^{\beta}\right) \tag{17}$$

其中 ψ 是 Digamma 函数。

Step3　根据贝叶斯定理将每一个特征点分配到最邻近的簇。

在人体动作识别模型的最后利用改进后的 MG-GMMs 模型有效地对人体动作进行建模和分类。

2　实验分析

2.1　数据集和评估指标

实验的数据集来源于公开数据集 KTH,该数据集包含 4 个场景下 25 人的 6 类不同动作,总共包含 2391 个视频序列(原数据集中丢失一个视频序列文件,文件名为 person13_handclapping_d3)。4 个场景分别是室外、室内、尺度变化的室外和视频中个体的衣着穿戴变化的室外,因此该数据集包含了尺度变化、衣着变化和光照变化。视频序列包含了 6 类动作分别是拳击(Boxing)、拍手(Hand clapping)、挥手(Hand waving)、慢跑(Jogging)、跑步(Running)、行走(Walking),KTH 数据集的一些示例图像如图 2 所示。

实验中,对数据集随机选取 80% 作为人体动作识别模型的训练集,20% 为测试集。将 KTH 数据集的人体动作的识别准确率作为实验的评估指标。

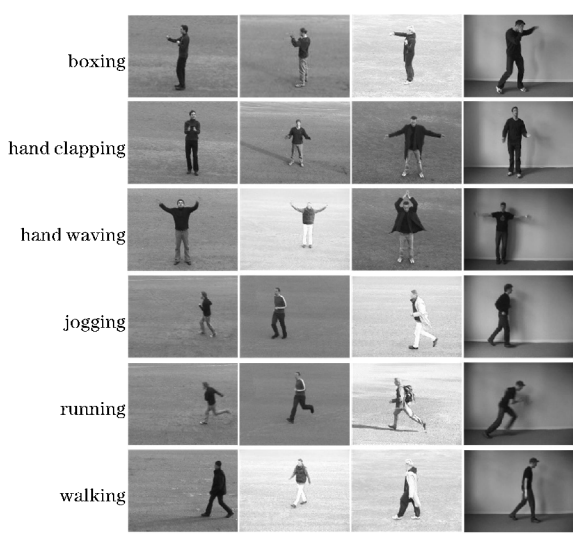


图 2　KTH 数据集 6 类动作示例图像

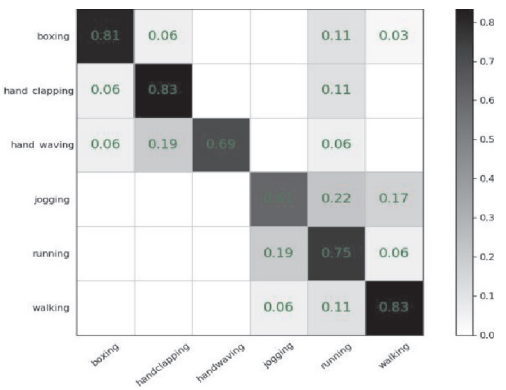
2.2　实验结果与分析

在 Ubuntu16.04 系统搭建的包含开源计算机视觉库(Opencv3)的 Pycharm 环境下进行实验。利用描述的人体动作识别模型对数据集进行识别分类,通过模型的第一步和第二步来检测和提取视频序列的兴趣点和描述子,然后通过第三步构建视觉词袋,最后利用改

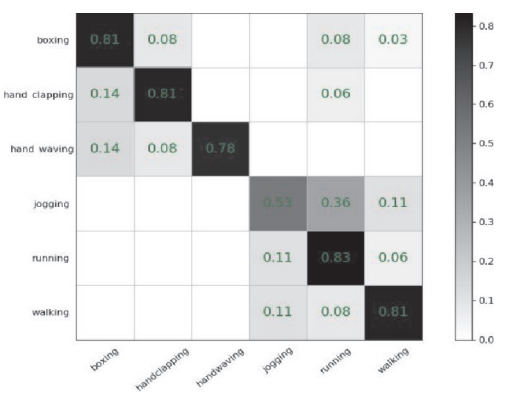
进后的 MGGMMs 模型对动作进行建模和分类。在第三步构建视觉词袋中,考虑到不同标准的聚类中心的数目对识别准确率的影响,对 200,500,600 这 3 个标准进行了对比实验,实验结果如表 1 所示,混淆矩阵如图 3 所示。

表 1 BOW 模型中不同 k 值的比较

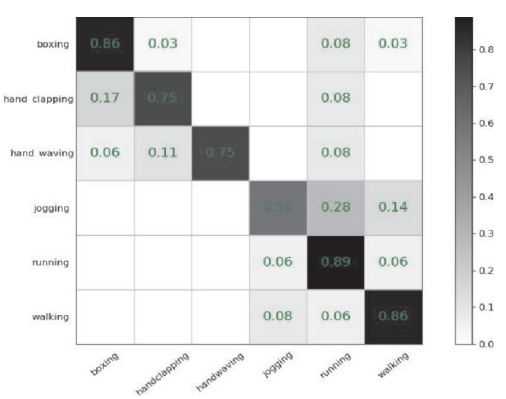
k	200	500	600
识别准确率/%	78.24	79.53	75.46



(a) k=200 时的混淆矩阵



(b) k=500 时的混淆矩阵



(c) k=600 时的混淆矩阵

图 3 BOW 模型中不同 k 值的混淆矩阵图示

根据表 1 不难看出,在构建视觉词袋的步骤中,当 k=500 时,识别准确率最高,因此在后续实验中,均采用 k=500 的参数设置。图 3 是对应 k 取值的混淆矩阵,x 轴为预测值。由图 3 还可以看出,识别准确率的短板主要在“jogging”,容易将其误判为“running”或“walking”。

表 2 展示了在识别模型中使用传统广义高斯混合模型(GGMMs)和使用的结合定点估计的多元广义高斯混合模型(MGGMMs)的识别准确率结果。由表 2 可以看出,使用的结合定点估计的多元广义高斯混合模型的识别准确率高于传统的广义高斯混合模型,证明了文中模型在人体动作识别任务上的有效性。

表 2 不同高斯混合模型的识别准确率比较

模型	识别准确率/%
GGMMs	83.56
文中模型	88.43

将方法与其他不同算法在 KTH 数据集上进行识别准确率比较,对比结果如表 3 所示。文中方法在 KTH 数据集上的最终人体动作识别准确率的混淆矩阵如图 4 所示。根据表 3 可以看出,相较于表中其他方法,文中方法在识别准确率上都有不同程度的提高,证明所提出的融合时空兴趣点和多元广义高斯混合模型的人体动作识别方法能够有效地对人体动作进行识别分类。

表 3 不同方法在数据集 KTH 上的识别准确率结果比较

方法	识别准确率%
文献[11]	81.17
文献[17]	83.33
文中模型	88.43

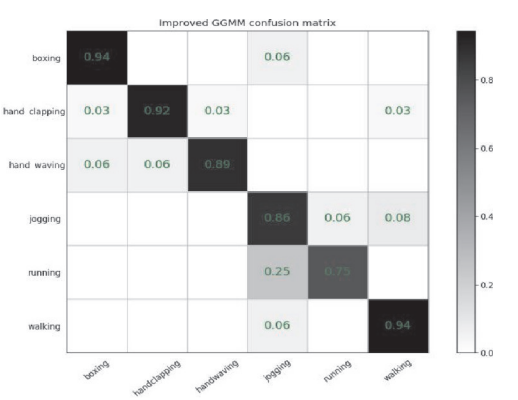


图 4 文中模型的人体动作识别准确率的混淆矩阵图示

3 结束语

根据实验结果和分析可以看出,融合时空兴趣点和多元广义高斯混合模型的人体动作识别方法在 KTH 数据集上相较于传统广义高斯混合模型取得了

较好的识别结果。该方法对于实验环境和硬件基本没有要求,在 Windows 系统和 Linux 系统下均能运行,在没有 GPU 硬件支持的情况下,也能在较短的时间内完成运行获得识别结果,对于小数据集较为友好。同时,由于该方法参数量少,计算成本较低,但距离实时应用还有一定的距离。而且由于词袋模型不容易考虑到语序问题,而视频序列中的人体动作包含了丰富的时间序列关系,因此在一定程度上在该步骤不能充分利用时空关系而获得更为准确的识别结果,只能依赖尽可能地提取到有效的时空兴趣点。同时,文中方法对于复杂环境下的多视角人体动作的识别具有一定的局限性。因此,后续对基于视频的人体动作识别研究可以在充分利用时空信息的前提下深入研究特征提取和特征表示,以获得更好的识别效果。

参考文献:

[1] Aggarwal J K, Ryoo M S. Human Activity Analysis: A Review[J]. ACM Computing Surveys, 2011, 43(3): 1-43.

[2] 胡琼,秦磊,黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报,2013(12):2512-2524.

[3] Piyathilaka L, Kodagoda S. Gaussian Mixture Based HMM for Human Daily Activity Recognition Using 3D Skeleton Features[C]. 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), Melbourne, VIC, Australia, 2013: 567-572.

[4] Baxter R H, Robertson N M, Lane D M. Human behavior recognition in data-scarce domains[J]. Pattern Recognition, 2015, 48(8): 2377-2393.

[5] Zhou Z, Shi F, Wu W. Learning Spatial and Temporal Extents of Human Actions for Action Detection[J]. IEEE Transactions on Multimedia, 2015, 17(4): 512-525.

[6] Caquetá J M, Carmona EJ, Fernandez-Caballero A. A survey of video datasets for human action and activity recognition[J]. Computer Vision and Image Understanding, 2013, 117(6): 633-659.

[7] Bregonzio M, Shaogang G, Tao X. Recognizing action as clouds of space-time interest points[C]. 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1948-1955.

[8] Selmi M, Yacoubi M. E, Dorizzi B. On the sensitivity of spatial-temporal interest points to person identity[C]. 2012 IEEE Southwest Symposium on Image Analysis and Interpretation, 2012: 69-72.

[9] Bellamine I, Tairi H. Motion detection and tracking using space-time interest points[C]. 2013 ACS International Conference on Computer Systems and Applications (AICCSA), 2013: 1-7.

[10] Hendaoui R, Abdellaoui M, Douik A. Synthesis of spatio-temporal interest point detectors: Harris 3D, MO SIFT and SURF-MHI[C]. 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2014: 89-94.

[11] Dollar P. Behavior recognition via sparse spatial-temporal features[C]. 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005: 65-72.

[12] Wang J, Chen Z, Wu Y. Action recognition with multiscale spatial-temporal contexts[C]. CVPR 2011: 3185-3192.

[13] 丁松涛,曲仕茹. 基于改进时空兴趣点检测的人体行为识别算法[J]. 西北工业大学学报,2016,34(5):886-892.

[14] Turaga P. Machine Recognition of Human Activities: A Survey[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(11): 1473-1488.

[15] Shaily S, Mangat V. The Hidden Markov Model and its application to Human Activity Recognition[C]. 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), 2015: 1-4.

[16] Piyathilaka L, Kodagoda S. Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features[C]. 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), 2013: 567-572.

[17] Niebles J C, Wang H, FeiFei L. Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.

[18] Bruno B., Human motion modelling and recognition: A computational approach[C]. 2012 IEEE International Conference on Automation Science and Engineering (CASE), 2012: 156-161.

[19] Ali N M. Object classification and recognition u-

sing Bag-of-Words (Bow) model[C]. 2016 IEEE 12th International Colloquium on Signal Processing & Its Applications(CSPA),2016:216-220.

[20] Ghildiyal B, Singh A, Bhadauria H S. Image-based monument classification using bag-of-word architecture[C]. 2017 3rd International Conference on Advances in Computing Communication & Automation(ICACCA)(Fall),2017:1-5.

[21] Piyathilaka L, Kodagoda S. Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features[C]. Industrial Electronics & Applications. IEEE,2013.

Human Motion Recognition based on Spatiotemporal Interest Points and Multivariate Generalized Gaussian Mixture Models

HE Bingqian, WEI Wei, SONG Yanbei, GAO Lianxin, ZHANG Bin

(College of Computer Science and Technology, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: Human action recognition has been widely used in the fields of human-computer interaction and virtual reality, etc. Aiming at the problems of excessive redundancies and neglecting the correlation of image data in traditional human motion recognition algorithm, this paper proposed a human motion recognition method based on spatio-temporal interest points and multivariate generalized Gaussian mixture model combined with fixed-point estimation. Redundant feature points and multivariate generalized Gaussian mixture models are used to effectively extract feature points and make full use of data correlation. The Harris-Laplace algorithm and 3D-SIFT descriptor extracted the feature points of the video sequence, and the BOW model clustered the visual words. Finally, the improved multivariate generalized Gaussian mixture model modeled and classified. Experiments were performed on the KTH datasethe experimental results show that the proposed human motion recognition method can effectively recognize and classify human motion in video.

Keywords: action recognition; spatial and temporal interest points; Harris-Laplace; 3D-SIFT; feature extraction