

基于最大相关-最小冗余的动作识别算法^①

龚 静¹, 李英杰¹, 黄欣阳²



同类文章推荐

1. 湖南环境生物职业技术学院 公共基础课部, 湖南 衡阳 421005;

2. 南华大学 计算机学院, 湖南 衡阳 421001

摘要: 为了提高动作的识别精度与鲁棒性, 降低冗余特征, 提高算法效率, 设计了一种基于最大相关-最小冗余(Max-Correlation and Min-Redundancy, MCMR)的动作识别算法. 首先, 为了消除噪声影响, 减少计算成本, 利用符号聚集近似(SAX)技术将连续图像序列转换为离散符号; 其次, 为避免出现时间漂移问题, 利用动态时间归整(Dynamic Time Warping, DTW)来计算符号特征的距离, 提取符号序列的特征; 然后, 为了消除冗余的特征, 定义了一个特征权重, 根据权重对特征进行降序排列, 引入最大相关-最小冗余技术消除相关性弱的特征, 筛选出具有高相关性和低冗余的特征; 最后, 为了完成动作识别, 根据筛选出的特征, 利用 k -近邻(K-Nearest Neighbor, KNN)进行分类器学习. 结果表明: 与当前动作识别算法相比, 本文算法能够有效完成动作的识别与理解, 具有较高的识别率, 有效地降低了冗余特征, 提高了算法的效率和鲁棒性.

关键词: 动作识别; 最大相关-最小冗余; 符号表示; 动态时间归整; k -近邻

中图分类号: TP399

文献标志码: A

文章编号: 1673-9868(2018)06-0158-10

动作识别是计算机视觉领域研究的热点问题, 在人机交互、监控、体感游戏、智能家居等方面具有重要的实用价值^[1]. 现如今, 视频已成为人们生活中不可缺少的一种信息载体, 随着视频呈现爆炸式地增加, 如何迅速、准确地提取视频中人们所关注的内容, 理解与识别图像序列中人体的动作意义重大^[2-3]. 近年来, 通过不断研究, 产生了一些关于动作识别的算法. 如杨顺卿等^[4]提出了一种基于双语动态系统的行为识别方法, 结合兴趣点检测与密集采样获取序列中的时空信息同时构建线性动态系统(LDS), 然后对 LDS 进行编码; 并利用 LDS 在码本中的分布与权重, 通过动态系统袋(bag of dynamical systems, BODS)表征动作特征; 接着通过 K-SVD 对得到的 BODS 特征操作, 形成迁移字典对; 最后通过 OMP 将字典在 2 个视角下进行动作的稀疏表示. 该算法能够连续完整地表达图像序列, 但是通过密集采样法会带来较多的冗余特征, 导致计算效率低下. 王鑫等^[5]提出了一种利用流形学习的动作识别技术, 该算法通过 Kinect 的深度内容获取活动位置, 并利用相对活动位移差来得到运动特征; 同时, 借助 Laplacian 映射流形对多维信号降维, 得到了低维运动信息, 在匹配过程中, 通过改进的 Hausdorff 对低维空间样本和训练数据的进行相似测量, 实验显示该算法识别效果良好. 但是, 此技术对视频中的视觉变化、相互遮挡等较为敏感. Davis 等^[6]提出了一种基于运动能量图的动作识别方案, 首先, 通过马氏距离获取模板间的距离, 然后, 根据观

① 收稿日期: 2017-06-14

基金项目: 国家自然科学基金项目(61300234); 湖南省教育厅项目(12C1056, 17C0566).

作者简介: 龚 静(1972-), 女, 硕士, 教授, 主要从事图像信息处理、数据挖掘、模式识别方面的研究.

通信作者: 李英杰.

测与训练间的距离,引入了 k -近邻完成对动作学习.该方法计算简单,但在选择时间间隔时不好确定.当间隔小,样本与训练的差别较小时,动作识别精度良好;相反,识别率不高,一般适用于简单动作识别.

为了降低动作识别中的特征冗余,提高识别精度和计算效率,本文提出了一种基于最大相关-最小冗余(MCMR)的动作识别方案.首先,在距离估计之前,使用 SAX 将时间序列转换为较短的符号序列,消除噪声影响的同时加快计算,并利用一个符号表示快速 DTW 测量.由于不同传感器种类与不同维度数据上动作的识别具有差异性,本文采用 MCMR 策略来选择合适的候选特征;然后,根据选择的特征,利用 k -近邻分类学习;最后,对所提动作识别技术进行测试.

1 动态时间归整

动态时间归整(DTW)能够解决优化问题,利用时间规整函数 $W(n)$ 表示测试模板与参考模板的时序关联,以获取 2 个模板匹配的最小累加距离所对应的规整方程^[7].

设 2 个时序 Q, C , 对应的长度为 n, m , 则

$$\begin{cases} Q = [q_1, q_2, \dots, q_n] \\ C = [c_1, c_2, \dots, c_m] \end{cases} \quad (1)$$

时序 Q, C 的距离设为 D , 表示为

$$D = \begin{bmatrix} d(q_1, c_1) & d(q_1, c_2) & \dots & d(q_1, c_m) \\ d(q_2, c_1) & d(q_2, c_2) & \dots & d(q_2, c_m) \\ \vdots & \vdots & & \vdots \\ d(q_n, c_1) & d(q_n, c_2) & \dots & d(q_n, c_m) \end{bmatrix} \quad (2)$$

如果 $n = m$, 直接计算时序 Q, C 的距离;如果 $n \neq m$, 需要进行对齐操作.而线性缩放是将短时序伸长到与长时序相同的长度,或把长时序缩短使其与短时序长度相等,然后进行比较.

为了对齐时序 Q, C , 根据 Q 与 C 的距离, 建立一个矩阵网格. 矩阵中的单元为 q_i 与 c_j 的距离 $d(q_i, c_j)$, $d(q_i, c_j)$ 越小, 相似性越高. 为了得到一条通过网格中不同点的路线, 使其经过的点为 2 个时序的对齐点. 如图 1 所示.

在本文算法中, 把要寻找的这条路线定义为规整路线 W . 为了寻找一条最好的 W , 来建立 Q, C 的映射, 令

$$W = w_1, w_2, \dots, w_k \quad (3)$$

其中, $w_k = (i, j)_k$ 为其第 k 单元.

W 需满足以下几个条件:

1) 边界条件: $w_1 = (1, 1)$, $w_k = (n, m)$;

2) 连续性: 设 $w_{k-1} = (a', b')$, 那么下个点 $w_k = (a, b)$ 符合 $(a - a') \leq 1$, $(b - b') \leq 1$, 表示无法越过某点来匹配, 只允许与自己相邻点对齐. 确保 Q, C 的所有坐标出现在 W 中;

3) 单调性: 对于 $w_{k-1} = (a', b')$, 下一个点 $w_k = (a, b)$, 需符合 $0 \leq (a - a')$, $0 \leq (b - b')$. 从而确保 W 中的点可以随着时间单调完成.

通过连续性与单调性约束, 每个点的路线就只有 3 个方向, 如果路线通过了点 (i, j) , 下个点可能为 $(i + 1, j)$, $(i, j + 1)$, $(i + 1, j + 1)$ 这 3 个点中的一个.

为了获得感兴趣的点, 通过 $W(n)$ 计算规整成本最优的路线:

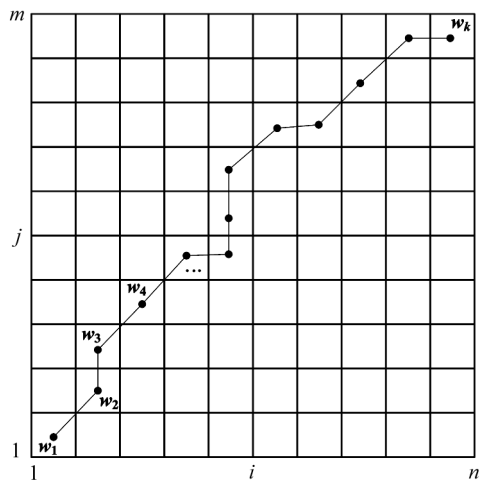


图 1 规整路径计算过程

$$DWT(Q, C) = \min \left\{ \sqrt{\frac{\sum_{k=1}^K \tau \omega_k}{K}} \right\} \quad (4)$$

通过将 2 个时序进行缩放, 可以获得 2 个时序的最短距离. 在此, 需选择一个路线, 使最终总距离最优. 为此, 本文通过一个累加距离, 将网格中所有点的距离进行相加, 形成最后的总距离, 因此, 需要计算时序 Q, C 的相似度, 计算公式如下所示:

$$\gamma(i, j) = d(q_i, c_i) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \quad (5)$$

最后, 通过动态规划(dynamic programming) 算法计算得到沿路线的累加距离, 此距离为最小值, 这条路线为最优路线.

2 算法设计

为了提高各种复杂动作的识别精度, 将每一个运动传感器的数据定义成不可重复的多元时间序列片段, 对于这种数据, 通常采用基于距离相似的方法来处理. 但是, 由于传感器种类的多样性和匹配效率存在多维性, 识别难度大, 通常采用 Shapelet 变换来提高时间序列的识别敏感性^[5]. 然而, 该方法存在数据维度高、时间消耗大的问题. 为此, 本文设计了一个基于符号表示的快速 DTW 测量方法, 在距离估计之前, 使用 SAX 将原来的时间序列转化为较短的符号序列, 以消除噪声影响, 提高计算效率. 因为传感器种类不同, 在不同维度数据上动作的识别具有差异性, 所以, 采用基于最大相关-最小冗余的策略来选择最佳的候选特征, 再根据选择的特征, 利用 k -近邻分类学习. 算法流程图见图 2.

基于距离度量的运动时间序列识别方法是将时间序列之间的距离作为一种相似性度量, 此方法在很多情况下都是有效的, 如测试样本和选定的图案之间的神经网络嵌入距离与相似性测量方面. 本文采用了测试运动系列与不同分类特征值序列之间的距离, 并将其应用于不同种类的分类学习中. 但是, 由于存在不同人体执行的相同动作或者同一人执行不同动作的情况, 因此, 本文采用 DTW 方法进行距离估计. 同时, 考虑到时间成本, 通过符号表示来降低数据维度.

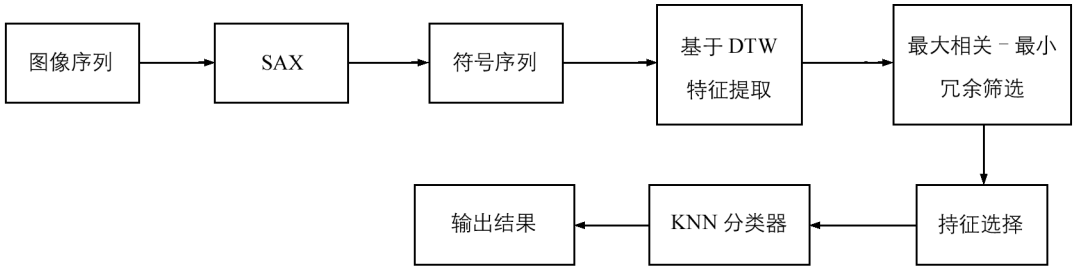


图 2 本文算法流程图

2.1 符号表示

设 $s = \{s_1, s_2, \dots, s_n\}$ 为多维时间序列, 表示一个单一的运动与维度传感器数据. $s_i = \{sd_1, sd_2, \dots, sd_m\}$ 为第 i 维度下长度为 m 的数据. 由于动作范围存在差异, 相似形状的不同运动也可能存在不同, 因此, 需要对数据标准化, 以消除这种分歧. 对于第 i 维的时间序列 s_i , 其标准化表示为:

$$s_{i-nor} = \frac{s_i - \mu_{si}}{\sigma_{si}} \quad (6)$$

其中, μ_{si} 为均值; σ_{si} 为标准差, 分别定义如下:

$$\mu_{si} = \frac{\sum sd_i}{m}, \sigma_{si} = \sqrt{\frac{\sum_{j=1}^m (sd_j - \mu_{si})^2}{m}} \quad (7)$$

为了消除噪声影响的同时减少时间消耗, 利用 SAX 将连续时间序列转换为离散符号序列^[8]. 通过分段聚集近似(Piecewise Aggregate Approximation, PAA) 技术, 将时间序列划分成相同大小的片段. 在相同

概率的高斯分布下,断点 $B(B_1, B_2, \dots, B_p)$ 将数据范围划分为 $p+1$ 段,并将所有的片段转化为用相应的符号表示. 对于一个时间序列 s_i , 设断点数为 p , 片段数为 f , 可得到一个符号序列 $st_i \{st_{i_1}, st_{i_2}, \dots, st_{i_f}\}$, $f < m$, 对于 st_i 中的每个元素, 满足 $st_{i_j} \in O_1, O_2, \dots, O_{p+1}$, $1 \leq j \leq f$, O 为一种符号的表示. 操作过程见图 3.

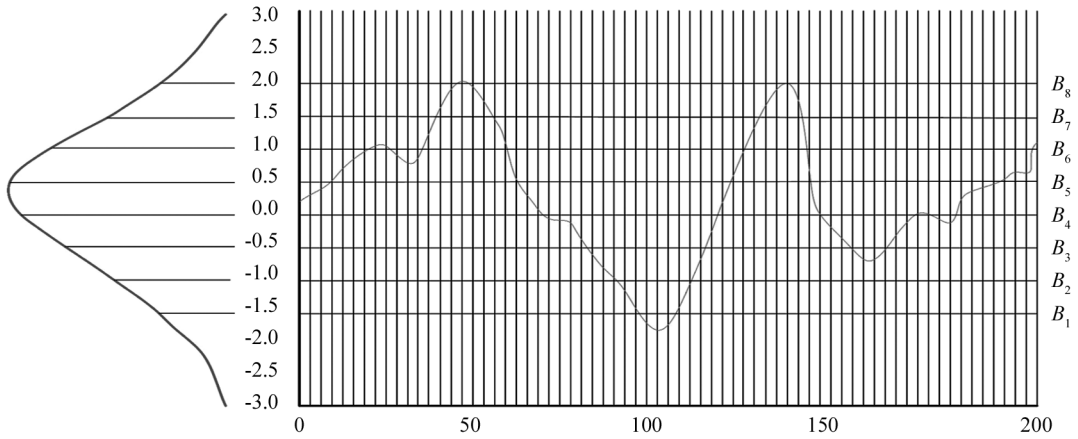


图 3 符号表示示意图

通过上述过程,能够完成尺寸缩减和离散化. 由于时间序列的长度变化会影响比较结果,因此,本文利用 SAX 将不同长度的运动转换为相同长度的序列.

2.2 基于 DTW 的特征提取

DTW 是时间序列相似估计中一种有效的距离度量手段,利用 SAX 得到的欧氏距离作为离散符号数据的下限距离^[9]. 假设 2 个符号序列 $st_x \{x_1, x_2, \dots, x_i, \dots, x_{ld}\}$, $st_y \{y_1, y_2, \dots, y_i, \dots, y_{ld}\}$, 它们的下限距离定义如下所示:

$$D_{\text{lower}}(st_x, st_y) = \sqrt{\frac{m_x + m_y}{2 * low}} \sqrt{\sum_{i=1}^{low} \psi(x_i, y_i)^2} \quad (8)$$

其中, m_x, m_y 分别为数据 x, y 的长度; low 为数据维数, $\psi(x_i, y_i)$ 计算如下所示:

$$\psi(x_i, y_i) = \begin{cases} 0 & |x_i - y_i| \leq 1 \\ B_{\max(x_i, y_i)-1} - B_{\max(x_i, y_i)} & \text{otherwise} \end{cases} \quad (9)$$

对于多维数据而言,传统的距离度量可提供简单和快速的相似度测量. 但这些距离序列对应 2 个时间序列,不适用偏移和时间漂移的场合,因此,本文引入了 DTW 来解决此问题. 同时,为了加快计算,本文利用 Sakoe-Chiba 自带限制的匹配范围^[10]. 对于符号序列 $st_x \{x_1, x_2, \dots, x_i, \dots, x_{ld}\}$, $st_y \{y_1, y_2, \dots, y_i, \dots, y_{ld}\}$, 其距离计算如下所示:

$$D(i, j) = \begin{cases} 0 & i = 0, j = 0 \\ d(i, j) + D(i-1, j) & j = 0, 0 < i < bw \\ d(i, j) + D(i, j-1) & i = 0, 0 < j < bw \\ d(i, j) + \min(D(i-1, j-1), D(i, j-1), D(i-1, j)) & i > 0, j > 0, i-j < bw \end{cases} \quad (10)$$

其中, bw 为带宽约束.

一般地,可利用 Euclidean 距离测量匹配过程中 x_i 与 y_i 之间的相似性,

$$d(i, j) = \sqrt{\sum_{k=1}^{dim} (x_{ik} - y_{jk})^2} \quad (11)$$

其中, dim 为数据点的维度; x_{ik}, y_{jk} 分别表示 x_i 与 y_i 的第 k 维数据. 匹配过程如图 4 所示.

对于三维时间序列,数据点在每个时间上是一个三维数据,通过不同维度的相互关系选择一次匹配或

在每个维度上的匹配. 这 2 种选择可以得到不同的结果, 分别设为 DTW_D, DTW_I . 事实上, DTW_I 比 DTW_D 具有更高的运动感知性. 虽然, 无法实现对 DTW_I 的不同距离累加, 但可作为特征候选之一, 通过相互间的最小相关性寻找到最有效的候选特征.

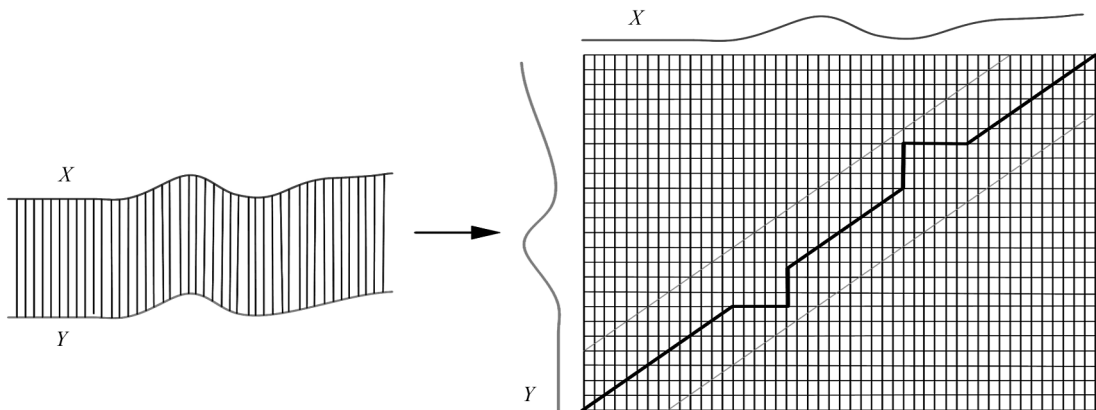


图 4 DTW 提取示意图

2.3 基于最大相关-最小冗余筛选的特征选择

提高动作识别的精度要从多方面来考虑, 其中特征选择是最最重要的一个方面. MCMR 能够筛选出具有不同类别、相互独立的特征^[11], 其选择过程主要由 2 个阶段组成: 第一阶段, 利用特征选择算法 ReliefF(RF) 来评估每个特征的区分能力^[12], 根据特征与类别的关联性来计算特征的权值, 权值小于阈值 τ 的特征将被删除, 选择出具有强类别相关性的特征作为候选特征; 第二阶段, 判断候选特征之间的相关性, 选择相关性低的特征, 直到收集了足够数量的特征为止.

ReliefF 算法适用于多类问题的特征选择, 通过几次迭代可确定每个特征的权重. 在每次迭代中, ReliefF 算法均是从标号数据集随机选择一个样本 R , 并从相同类别中提取 k 个最近邻样本. 随后, 计算其特征权重:

$$RF(F_i) = w(F_i) - \sum_{j=1}^k \frac{ff(F_i, R, H_j)}{nk} + \sum \frac{[(p(C)/(1-p(R))) \sum_{j=1}^k ff(F_i, R, M_j(C))]}{nk} \quad (12)$$

ff 函数定义如下:

$$ff(F_i, R_1, R_2) = \begin{cases} 0 & R_1(F_i) = R_2(F_i) \\ 1 & R_1(F_i) \neq R_2(F_i) \\ \frac{|R_1(F_i) - R_2(F_i)|}{\max(F_i) - \min(F_i)} & F_i \text{ 连续} \end{cases} \quad (13)$$

其中, F 是从不同类别的标记数据中选择的候选特征; $w(F_i)$ 为第 i 个候选特征的权重; H_j 为第 j 次近邻采样; $M_j(C)$ 为类别 C 的第 j 次近邻遗漏; $\max(F_j)$ 与 $\min(F_j)$ 为 F_j 的最大和最小值, $p(C)$, $p(R)$ 分别为类别 C 、样本 R 的概率.

但是, 在上述过程中, 候选特征忽略了特征间的关联和冗余, 在实际过程中, 需要消除冗余的候选特征. 因此, 根据 RF 权值, 对特征进行降序排列, 消除小于阈值的特征. 对于剩余的候选特征, 检查其是否具有相关性, 如果一个候选特征与所选择的其他特征有密切的关系, 也要剔除该特征. 否则, 将其作为最终的选择特征, 直到收集了足够数量的特征为止.

在以往的研究中, 提出了各种用于估计特征冗余的方法, 例如信息增益、互信息、相关系数^[13] 等. 但是, 冗余的候选特征通常有相同的类可分性, 意味着它们的 RF 权重具有较高的线性相关性. 因此, 本文采

用基于 RF 权值的 Pearson 相关系数来表示特征间的关联^[14]:

$$P(a,b)=\frac{\sum w_a w_b-\sum w_a \sum w_b/N}{\sqrt{(\sum w_a^2-(\sum w_a)^2/N)(\sum w_b^2-(\sum w_b)^2/N)}}$$

(14)

其中, w_a 表示所选特征的 RF 权重矩阵; w_b 表示候选的 RF 权重矩阵; $P(a,b)$ 表示 Pearson 相关系数. Pearson 相关系数主要用于计算 2 个变量间的线性相关程度, 其值是在 -1 和 1 之间. 1 表示呈正相关; 0 表示无相关性; -1 表示呈负相关. 如果这个相关系数高于阈值, 那么淘汰该候选特征.

2.4 分类器学习

为了准确地完成动作识别, 需要对所选择的特征进行分类学习. 本文引入了 k -近邻(KNN)技术^[15]. KNN 是一种简单的非参数学习算法, 对给定的数据中, 搜索其最靠近的 k 个临近点, 临近点最多的种类为该样本类别. k -近邻在分类时, 只需利用最近的几个类别来确定未知类别, 通过少量的样本就可避免分类不平衡问题.

令特征为 $\{a_1(x), a_2(x), \cdots, a_n(x)\}$, $a_i(x)$ 为 x 的第 i 个特征值, 则样本 X_M 与 X_N 的距离表示为

$$D(X_M,X_N)=\sqrt{\sum_{i=1}^n(a_i(X_M)-a_i(X_N))^2}$$

(15)

KNN 的计算过程如下所示:

- 1) 对特征归一化处理;
- 2) 计算已知点和求解点的距离;
- 3) 按照距离递增规律进行排列;
- 4) 选择与当前点距离最近的 k 个点;
- 5) 计算前 k 个点的概率;
- 6) 对前 k 个点中概率最大的特征进行类别召回, 并判断为类别对象.

图 5 显示了 KNN 分类操作. 为了对未知的圆分类, 令样本 $k=3$, 通过分析计算, 圆和正方形接近数量高于三角形. 因此, 判定圆为正方形类别的.

基于近邻加权的 KNN 能够很好地降低噪声干扰, 提高鲁棒性, 并且计算精度高、对异常值不敏感.

3 实验测试与分析

为了验证所提动作识别算法的有效性, 借助 Kinect 与 IXMAS 数据集来评价算法的性能. 开发工具: Linux 系统, QtCreator + OpenCV. 测试条件为: CoreI 3, 3.30GHz CPU, 8GB 运行 RAM. 为体现算法的有效性与先进性, 将本文算法与文献[4]算法、文献[5]算法和文献[6]算法进行比较. 经过多次试验, 参数设定如下: 断点数为 $p=5$, $\eta=0.5$, 数据维度为 50, k -近邻分类器中 $k=8$. 本文采用的数据集与方法见表 1 所示.

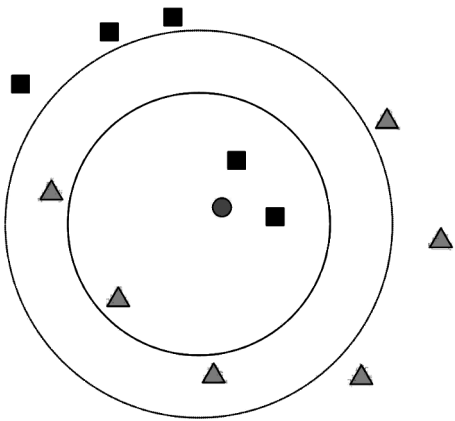


图 5 KNN 分类过程

表 1 实验数据集与方法

数据集	动作介绍	动作数量	实验方法
Kinect	走路/坐下/站起来/拿起/搬运/投掷/推/拉/挥舞/拍手	10 个对象, 200 个动作	10 倍交叉法
IXMAS	Check watch/cross arms/get up/kick/pick up/punch/scratch head/sit down/turn around/walk/wave	11 个对象, 396 个动作	50%训练/50%测试

3.1 数据集

Kinect 数据集是通过 Kinect 设备拍摄所形成的, 通过 10 个对象在 10 种不同动作下重复 2 次从而形成了 200 个序列的运动集. Kinect 数据集^[16]中的动作包括走路、坐下、站起来、拿起、搬运、投掷、推、拉、挥舞、拍手, 见图 6. 数据集提供深度和骨架信息, 对于这个数据集, 遵循了跨学科的设置. IXMAS 数据集^[17]含有 11 种动作, 每个动作分别由 12 个人执行 3 次, 并且从 5 个不同方向进行拍摄, 从而形成的一个多视角的数据集. 11 种动作包括: Check watch, cross arms, get up, kick, pick up, punch, scratch head, sit down, turn around, walk, wave. 该数据集包含了手臂、腿、躯干等相关部位中较全面的动作类型, 具有较强的说服力.



图 6 Kinect 运动集

3.2 实验结果

表 2 显示了本文算法在 Kinect 数据集中的混淆矩阵. 根据表中的测量结果可知, 本文算法在 Kinect 数据集中具有优异的识别能力, 对各动作的识别率较高, 能够对大多数动作进行准确识别. 如“站起来”、“投掷”、“挥舞”等动作的识别率高达 1. 但有个别行动识别效果不太理想, 如“走”、“推”动作识别率只有 0.85, “走”易被作为“搬运”动作, “推”易被误认为“投掷”动作. 主要是因为这些动作具有较多的相似性, 导致其不能有效区分其提取的动作特征.

表 2 Kinect 数据集混淆矩阵

	走	坐下	站起来	拿起	搬运	投掷	推	拉	挥舞	拍手
走	<u>0.85</u>		0.03		0.10	0.02				
坐下		<u>0.95</u>	0.05							
站起来			<u>1.00</u>							
拿起		0.05		<u>0.90</u>				0.05		
搬运		0.01	0.08		<u>0.90</u>	0.01				
投掷						<u>1.00</u>				
推				0.02		0.08	<u>0.85</u>		0.05	
拉		0.02		0.03				<u>0.95</u>		
挥舞									<u>1.00</u>	
拍手							0.03		0.02	<u>0.95</u>

表 3 为本文算法在 IXMAS 数据集中得到的混淆矩阵. 从表 3 中可看出, 本文算法在 IXMAS 数据集中的识别精度较为理想, 对各动作的识别率较高, 能够对大多数动作进行准确识别. 如“kick”, “scratch head”, “walk ”等动作的识别率高达 1. 但有个别行动识别效果不太理想, 如“pick up”动作识别率只有 0.80, “sit down”动作识别率只有 0.85. 因为“pick up”易被误认为“get up”、“sit down”动作, “sit down”易被误认为“get up”、“pick up”动作, 这些动作具有较多的相似性, 区别起来难度较大.

表 3 IXMAS 数据集混淆矩阵

	Check	cross	get	kick	pick	punch	scratch	sit	turn	walk	wave
	watch	arms	up		up		head	down	around		
Check watch	<u>0.90</u>	0.05									0.05
cross arms		<u>0.95</u>	0.05								
get up			<u>0.90</u>		0.05			0.05			
kick				<u>1.00</u>							
pick up			0.08		<u>0.80</u>	0.05		0.07			
punch						<u>0.95</u>					0.05
scratch head							<u>1.00</u>				
sit down			0.06	0.02	0.07			<u>0.85</u>			
turn around						0.05			<u>0.95</u>		
walk										<u>1.00</u>	
wave		0.05				0.05					<u>0.90</u>

为了获得最优的性能，降低算法的复杂度，在不同的离散分割目标维数下测量算法的识别精度，结果见图 7。从图 7 中可知，在目标维数为 50 时本文算法的识别性能最优。

在 Kinect 与 IXMAS 数据集中，测试了 3 个对照算法以及本文算法的平均识别精度，如图 8 所示。依据图 8 可知，相对对照算法，本文算法在 Kinect 与 IXMAS 数据集中的识别率取得了优异的成绩，在 Kinect 与 IXMAS 中的平均识别率分别为 96.23%，93.36%，与对照组算法相比有一定的提高。主要是因为本文利用 SAX 技术将连续的图像序列转换为离散符号序列，消除噪声影响的同时也减少了时间消耗；同时引入 DTW 计算符号序列的距离，提取时间序列的特征，解决时间漂移问题；然后，定义了一个 *RF* 特征权重，根据权重对特征进行降序排列，消除相关性弱的冗余特征，筛选出具有高相关性和低冗余的特征；最后，为了完成动作识别，根据筛选出的特征进行分类器学习，降低噪声干扰，提高了算法鲁棒性。而文献[4]是通过密集采样得到特征数据的，容易带来较多的冗余特征；文献[5]是基于位移差来得到运动特征的，容易受视频中的视觉变化、相互遮挡等动作的影响，识别效果不好；文献[6]在选择时间间隔时不好确定，一般用于简单动作识别中。

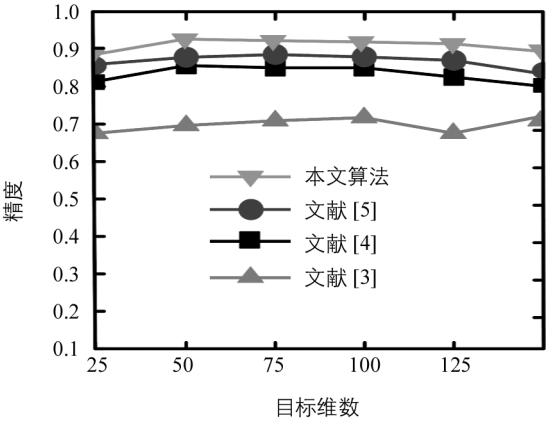


图 7 不同目标维数下的性能比较

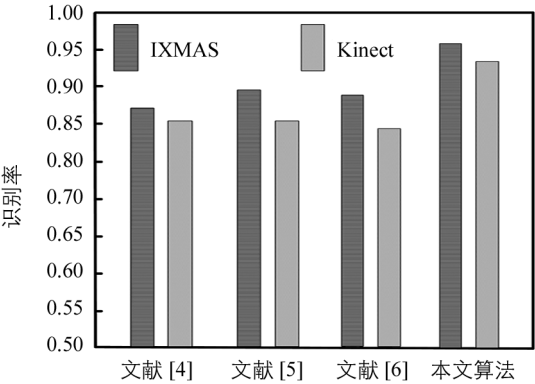


图 8 不同算法的识别正确率测试

为了更全面检测算法的性能，对算法的运行时间进行测试，结果见表 4。根据表 4 可知，本文算法在 Kinect 与 IXMAS 数据集中的平均运行时间为 31.25 s, 51.64 s，均低于文中的对照组算法，这表明本文算法具有较高的运算速度。其主要原因是本文利用了 SAX 方法将连续的图像序列转换为较短的离散符号序列，同时通过 MCMR 策略来选择合适的候选特征，消除冗余特征，有效地提高了算法效率。

表 4 算法效率评估

算法名称	运 行 时 间/s	
	Kinect	IXMAS
文献[4]	183.95	301.66
文献[5]	54.87	98.20
文献[6]	33.58	73.21
本文算法	31.25	51.64

4 结 论

为了提高动作的识别精度与鲁棒性,降低冗余特征,提高算法效率,本文提出了一种基于最大相关-最小冗余的动作识别算法.为了消除噪声影响并提高算法的效率,在距离估计之前,通过使用 SAX 技术将原来的时间序列转换为较短的符号序列;采用 DTW 准确提取图像时间序列中的特征,避免了时间漂移;针对动作特征的相似性和差异性,采用了 MCMR 策略来选择合适的候选特征,MCMR 根据特征权重对特征进行降序排列,消除弱相关性特征,选择高相关与低冗余特征;再根据选择的特征,利用 k -近邻分类学习.实验证明了本文算法具有优异的识别率,降低了冗余特征,并有效地改善了算法的识别效率.本文的不足之处在于只在 2 个数据集上进行了实验,未来将着眼于处理更多的数据集,还将研究基于深度神经网络的人体动作识别.

参考文献:

[1] 陈文强,肖国强,林 霄,等.一种基于属性贝叶斯网络的行为识别模型 [J]. 西南师范大学学报(自然科学版),2014,39(3):7—11.

[2] DERPANIS K G, SIZINTSEV M, CANNONS K J, et al. Action Spotting and Recognition Based on a Spatiotemporal Orientation Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 527—540.

[3] 田国会,尹建芹,韩 旭,等.一种基于关节点信息的人体行为识别新方法 [J]. 机器人,2014,36(3):285—292.

[4] 杨顺卿,陈昌红.基于双语动态系统包的视角无关的人体行为识别 [J]. 南京邮电大学学报(自然科学版),2014,34(1):103—110.

[5] 王 鑫,沃波海,管 秋,等.基于流形学习的人体动作识别 [J]. 中国图象图形学报,2014,19(6):914—923.

[6] DAVIS J W, BOBICK A F. The Representation and Recognition of Human Movement Using Temporal Templates [C]// Computer Vision and Pattern Recognition, 1997. Proceedings. , 1997 IEEE Computer Society Conference on. San Juan, Puerto Rico, USA: IEEE, 1997: 928—934.

[7] 傅 颖,郭晶云.基于动态时间规整的人体动作识别方法 [J]. 电子测量技术,2014,37(3):69—72.

[8] 宋 伟,张 帆,叶阳东,等.面向时间序列分类任务的 SAX 方法特性及可视化探索 [J]. 计算机辅助设计与图形学学报,2016,28(11):1906—1915.

[9] WEI Y, JIAO L B, WANG S L, et al. Sports Motion Recognition Using MCMR Features Based on Interclass Symbolic Distance [J]. International Journal of Distributed Sensor Networks, 2016, 12(5): 1—15.

[10] LU S, MIRCHEVSKA G, PHATAK S S, et al. Dynamic Time Warping Assessment of High-Resolution Melt Curves Provides a Robust Metric for Fungal Identification [J]. PLoS ONE, 2017, 12(3): e0173320.

[11] 姚明海,王 娜,齐 妙,等.改进的最大相关最小冗余特征选择方法研究 [J]. 计算机工程与应用,2014,50(9):116—122.

[12] ROBNIK-SIKONJA M, KONONENKO I. Theoretical and Empirical Analysis of ReliefF and RReliefF [J]. Machine Learning, 2013, 53(1): 23—69.

[13] 李 扬,顾雪平.基于改进最大相关最小冗余判据的暂态稳定评估特征选择 [J]. 中国电机工程学报,2013,33(34):179—186.

- [14] JOHNSON M, BRIGG M, GRAHAM M. Pearson and Responsibility: (Mis-) Understanding the Capabilities Approach [J]. *Australian Journal of Politics & History*, 2016, 62(2): 251–267.
- [15] 朱郭峰, 杨彦, 周竹荣, 等. 基于领域的微博用户影响力计算方法 [J]. *西南大学学报(自然科学版)*, 2014, 36(3): 145–151.
- [16] XIA L, CHEN C C, AGGARWAL J K. View Invariant Human Action Recognition Using Histograms of 3D Joints [C]//Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. Providence, RI, USA: IEEE, 2012: 20–27.
- [17] 姚湘, 徐平平, 王华君. 基于非线性知识迁移的交叉视角动作识别 [J]. *重庆邮电大学学报(自然科学版)*, 2017, 29(1): 121–128.

An Action Recognition Algorithm Based on Maximum Correlation-Minimum Redundancy

GONG Jing¹, LI Ying-jie¹, HUANG Xin-yang²

1. Department of Public Basic Course, Hunan Environment Biological Polytechnic, Hengyang, Hunan 421005, China;

2. College of Computer Science and Technology, University of South China, Hengyang, Hunan 421001, China

Abstract: In order to improve the accuracy and robustness of motion recognition, reduce redundant features and increase algorithm efficiency, an action recognition scheme based on MCMR (maximum correlation-minimum redundancy) is designed. Firstly, in order to eliminate the influence of noise and reduce computational cost, the continuous image sequence is transformed into discrete symbol representation by using the symbolic aggregation approximation technology. Next, to avoid the problem of time drift, dynamic time warping (DTW) is used to calculate symbol distance and to extract the character of a symbolic sequence. Then, in order to eliminate redundant features, a feature weight is defined, the features are arranged in a descending order based on weight, the maximum correlation-minimum redundancy is introduced to eliminate the weak correlation feature, and features with high correlation and low redundancy are selected. Finally, in order to perform action recognition, classifier learning is performed, using k -nearest neighbor based on the selected features. Test results show that compared with the current image action recognition, this algorithm can effectively accomplish action recognition and understanding. It has a high recognition rate, the redundant features are greatly reduced, and the efficiency and robustness of the algorithm are improved.

Key words: action recognition; maximum correlation-minimum redundancy; symbolic representation; dynamic time warping; k -nearest neighbor

责任编辑 崔玉洁

