

文章编号 1004-924X(2017)06-1669-10

采用自适应变异粒子群优化 SVM 的行为识别

张国梁*, 贾松敏, 张祥银, 徐 涛
(北京工业大学 信息学部, 北京 100124)

摘要: 为了提高对视频序列中人体行为的识别能力, 建立了基于局部特征的动作识别框架。通过时空特征提取及编码和 SVM 分类器参数优化两部分对该框架所涉及算法进行了研究。首先, 采用 Harris3D 检测器获取时空兴趣点 (STIP), 以方向梯度直方图 (HOG) 和光流方向直方图 (HOF) 对 STIP 进行描述, 并引入 Fisher 向量实现对特征描述子的编码; 由于固定参数下 SVM 动作分类模型存在泛化能力不足的问题, 将粒子群算法应用于各动作分类器参数寻优过程中, 针对种群多样性逐代变化的特点, 构建粒子聚集度模型, 并利用其动态调节各代粒子的变异概率; 最后, 利用 KTH 和 HMDB51 数据集对所提方法进行验证。结果表明, 所提自适应变异粒子群算法 (AMPSO) 能够有效避免种群陷入局部最优, 具备较强的全局寻优能力; 在 KTH 和 HMDB51 数据集上的识别准确率分别为 87.50% 和 26.41%, 优于其余 2 种识别方法。实验证明, AMPSO 算法收敛性能良好且整体识别框架具有较高的实用性和准确性。

关键词: 人体行为识别; 自适应变异粒子群算法; 时空兴趣点; 特征编码; 支持向量机

中图分类号: TP391 **文献标识码:** A **doi:** 10.3788/OPE.20172506.1669

Action recognition based on adaptive mutation particle swarm optimization for SVM

ZHANG Guo-liang*, JIA Song-min, ZHANG Xiang-yin, XU Tao

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

* Corresponding author, E-mail: 285719262@qq.com

Abstract: The action recognition framework based on local features was established to improve the recognition ability of human behavior in video sequences. The algorithms related to the framework were researched through spatial temporal features extracting and encoding and parameters optimization of SVM classifier. Firstly, the feature descriptors composed of Histograms of Oriented Gradients (HOG) and Histograms of Optical Flow (HOF) were used to describe Space Time Interest Points (STIP) achieved by the Harris 3D detector and then encoded by Fisher Vector (FV). Due to the generalization ability of Support Vector Machine (SVM) model for action classification under fixed parameters was insufficient, the particle swarm optimization algorithm was applied to the parameter optimization of each action classifier. According to the characteristics of population diversity changed from generation to generation, the constructed particles aggregation degree model was used to adjust mutation probability for each generation of particles dynamically. Finally, the proposed method was

收稿日期: 2016-12-23; **修订日期:** 2017-01-24.

基金项目: 国家自然科学基金项目 (No. 61175087); 北京工业大学智能机器人“大科研”推进计划“助老智能轮椅床自主测控系统的研究与实现”资助项目 (No. 040000546317552)

verified by KTH and HMDB51 data sets. The results show that the Adaptive Mutation Particle Swarm Optimization (AMPSO) algorithm can avoid the local optimum and has strong global optimization capability. The recognition accuracies on KTH and HMDB51 data sets are 87.50% and 26.41%, respectively, which are better than two recognition methods. Experimental results indicate that the AMPSO algorithm has good convergence performance and the overall recognition framework has high practicability and accuracy.

Key words: human action recognition; Adaptive Mutation Particle Swarm Optimization (AMPSO); Space Time Interest Points (STIP); feature coding; Support Vector Machine (SVM)

1 引 言

目前,基于视频的人体行为理解被广泛应用于机器人人机交互、视频监控、虚拟现实以及病患监护等诸多领域^[1]。然而,因其行为速度、拍摄尺度、视角和动作背景的复杂多变以及动作类别自身定义的不确定性,导致行为分析依然是一项具有挑战性的研究工作。

局部特征表达^[2]是通过提取时空变化显著的局部区域的底层特征,并将其编码为视频的中层语义描述。无需对人体进行分割,且不受复杂环境背景及遮挡的影响,受到了众多研究者的关注。Schuldt 等^[3]首先将视觉词袋模型(Bag of Words, BOW)应用于解决动作识别问题,并将识别过程归结为 6 个步骤,即局部特征提取、特征预处理、词典生成、特征编码、码字合并以及特征分类。在局部特征提取方面, Laptev 等^[4]提出的时空兴趣点(Space Time Interest Points, STIP)和 Wang 等^[5]提出并改进的稠密轨迹特征(Improved Dense Trajectories, IDT)在 UCF101、Hollywood2 等数据集上取得了良好的测试效果。特征编码作为该框架下的研究热点,近年来取得了较大的研究进展。Liu 等^[6]提出了一种局部软分配编码方法,并在多个数据集上验证了其有效性。Fisher 向量作为一种超向量编码方法由 Perronnin 等^[7]将其应用于大规模图像分类领域,并取得了优良的测试效果。

在特征分类方面, Vapnik 基于风险最小化理论提出的支持向量机(Support Vector Machine, SVM)是机器视觉识别任务中应用较为广泛的分类模型^[8],其分类精度很大程度上取决于模型自身的参数设置,如惩罚参数及核函数带宽。传统

的参数优化方法如网格搜索法和梯度下降法等,在寻优过程中容易陷入局部最优^[9],因此研究者将群体智能优化算法引入到 SVM 参数寻优过程中。Huang 等^[10]提出一种基于遗传算法的 SVM 参数优化方法,并有效提高了测试集的分类精度。Shen 等^[9]将果蝇算法应用在了 SVM 参数寻优问题中,在医学诊断分类方面其结果明显优于细菌觅食和遗传算法。Wang 等^[11]利用人工鱼群算法对 SVM 参数进行了优化用以解决回归问题,该其算法缺乏实际的应用背景,可扩展性有待进一步研究。总体而言,上述方法的执行步骤相对繁多,计算复杂度较高,使得全局寻优过程的计算成本过大,不利于在大规模视频动作数据集中应用。粒子群算法(Particle Swarm Optimization, PSO)以鸟类捕食过程为参照,具有严谨的数学模型及收敛性证明,并且计算复杂度相对较低^[12]。但是传统 PSO 在寻优过程中存在陷入局部最优和早熟收敛的问题,从而影响 SVM 最优参数的获取。为了进一步提升基于局部特征的动作识别准确率,本文首先对动作视频提取 STIP 特征,通过主成分分析^[13-14](Principal Component Analysis, PCA)对特征做降维处理,并采用 Fisher 向量对所提特征进行编码得到视频的中层表达。在分类阶段,为了提高 SVM 模型的判别性能,提出了一种自适应变异粒子群算法(Adaptive Mutation Particle Swarm Optimization, AMPSO)。该方法能够针对种群多样性逐代变化的特点,构建粒子聚集度模型,并用其动态调节各代粒子的变异概率。此外,为了解决多分类情况下训练参数单一导致的模型泛化能力减弱问题,对不同动作类别赋予相应的优化参数,进而得到具有较高判别性能的分类模型。最后,在 KTH 和 HMDB51 数据集上对所提方法的有效性进行了验证。

2 视频特征提取及编码

对视频段提取时空兴趣点,利用 PCA 对底层特征进行降维处理,构建高斯混合模型用于 Fisher 向量编码,经过求和合并^[2]以及归一化处理后级联得到动作视频的中层表述。

2.1 时空兴趣点提取

在传统的 Harris2D 角点检测基础上加入时间尺度,进而得到 Harris3D 检测方法^[4]。与 Harris2D 角点检测方法类似,首先利用公式(1)构筑图像序列 $I(x, y, t)$ 的 3D 尺度空间,

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

其中, x, y 和 t 用于表述空间像素点的位置坐标, σ^2 和 τ^2 分别代表空间和时间因子, g 为高斯核,可通过公式(2)进行计算。

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

计算像素时空尺度空间二阶矩阵 μ 的特征值,如式(3),并取公式(4)所示响应函数的局部最大值作为视频的 Harris3D 角点。 $L_x = \partial_x(g * I)$, $L_y = \partial_y(g * I)$, $L_t = \partial_t(g * I)$ 分别表示经过高斯核模糊后的图像序列 I 的一阶偏导数。

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

$$H = \lambda_1 \lambda_2 \lambda_3 - k(\lambda_1 + \lambda_2 + \lambda_3)^3, \quad (4)$$

其中, λ_1, λ_2 和 λ_3 为矩阵 μ 的特征值, k 一般取为 0.005^[4]。为了实现对时空兴趣点的描述,将其时空领域 $(\Delta_x, \Delta_y, \Delta_t)$ 分割为 $3 \times 3 \times 2$ 的网格,并在各个网格中分别求取 4 个梯度方向量化的 HOG 特征以及 5 个光流方向量化的 HOF 特征,进而得到视频段的底层特征描述集。动作视频提取 STIP 的可视化效果如图 1 所示。

2.2 基于 Fisher 向量的特征编码

传统的投票式编码策略仅能表达局部特征描述子与视觉词汇之间的从属关系,容易导致较大的量化误差。Fisher 则由于融合了生成式与判别式模型的优势,在表述视觉词汇出现频率的同时,可以对局部特征在各词汇的分布进行编码,从而



(a) Walking

(b) Climb stairs

图 1 视频时空兴趣点检测效果示意图

Fig. 1 Detection effect of STIP for video sequences

获取更加完备的图像信息。

首先通过训练获取高斯混合模型的参数集 $\lambda = \{\omega_k, \mu_k, \Sigma_k, k = 1, \dots, K\}$, $\omega_k, \mu_k, \Sigma_k$ 依次表示第 k 个高斯单元的混合权重、均值向量和协方差矩阵。设 $X = \{x_t \in R^D, t = 1 \dots T\}$ 为某一视频段提取的特征描述集,且各 x_t 之间满足独立同分布, D 是降维后的特征向量维度,则 x_t 的概率密度函数 P_λ 可由公式(5)表示。 P_k 为 x_t 的第 k 个高斯单元,如式(6)所示:

$$p_\lambda(\mathbf{x}_t) = \sum_{k=1}^K \omega_k p_k(\mathbf{x}_t). \quad (5)$$

$$p_k(\mathbf{x}_t) = \frac{\exp\left\{-\frac{1}{2}(\mathbf{x}_t - \mu_k)' \Sigma_k^{-1} (\mathbf{x}_t - \mu_k)\right\}}{(2\pi)^{D/2} |\Sigma_k|^{1/2}}. \quad (6)$$

假定协方差矩阵为对角阵,即 $\sigma_k^2 = \text{diag}(\Sigma_k)$ 。 $\gamma_t(k)$ 表示局部特征 x_t 分配给第 k 个高斯单元的概率,其计算过程可由公式(7)表述:

$$\gamma_t(k) = \frac{\omega_k p_k(\mathbf{x}_t)}{\sum_{i=1}^K \omega_i p_i(\mathbf{x}_t)}. \quad (7)$$

依据上述推导过程, X 关于 $\omega_k, \mu_k, \Sigma_k$ 的归一化梯度向量可由公式(8)~(10)计算得出,

$$G_{\omega_k}^X = \frac{1}{\sqrt{\omega_k}} \sum_{t=1}^T (\gamma_t(k) - \omega_k), \quad (8)$$

$$G_{\mu_k}^X = \frac{1}{\sqrt{\omega_k}} \sum_{t=1}^T \gamma_t(k) \left[\frac{x_t - \mu_k}{\sigma_k} \right], \quad (9)$$

$$G_{\sigma_k}^X = \frac{1}{\sqrt{2\omega_k}} \sum_{t=1}^T \gamma_t(k) \left[\frac{(x_t - \mu_k)^2}{\sigma_k^2} - 1 \right]. \quad (10)$$

利用 soft-max 机制对混合权重 ω_k 进行定义,如公式(11)所示。将 K 个高斯单元上的 $G_{\omega_k}^X$ 、 $G_{\mu_k}^X$ 和 $G_{\sigma_k}^X$ 进行级联,从而得到 $(2D+1)K$ 维的 Fisher 向量,表示为公式(12)。

$$\omega_k = \exp(\alpha_k) / \sum_{j=1}^K \exp(\alpha_j). \quad (11)$$

$$G_{\lambda}^X = (G_{\alpha_1}^X, \dots, G_{\alpha_k}^X, G_{\beta_1}^X, \dots, G_{\beta_k}^X, G_{\gamma_1}^X, \dots, G_{\gamma_k}^X)'. \quad (12)$$

在所得 Fisher 向量基础上对其每一维度执行能量归一化,并对整体做 L2 归一化操作^[15],进而得到视频的中层表达,如公式(13)和(14)所示。

$$[G_{\lambda}^X]_i \leftarrow \text{sign}([G_{\lambda}^X]_i) \sqrt{|[G_{\lambda}^X]_i|}, \quad (13)$$

$$G_{\lambda}^X = G_{\lambda}^X / \sqrt{G_{\lambda}^X{}^T G_{\lambda}^X}. \quad (14)$$

3 变异粒子群优化 SVM 分类

3.1 支持向量机分类

SVM 能够通过构造最优分类超平面,完成对测试样本的线性分类任务。设某二分类样本集为 $\{x_i, y_i\}$, x_i 是 D 维的样本点, $y_i \in \{-1, 1\}$ 表示 x_i 对应的标签, $i=1, \dots, l$ 。分类超平面可以表述为 $w^T x + b = 0$, 其中 w 是超平面的法向量, b 为相应的常数项, 则线性 SVM 可通过公式(15)获取最佳分离边缘。

$$\min g(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i. \quad (15)$$

$$s. t., y_i (w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0. \quad (16)$$

采用拉格朗日乘子 $\alpha_i (i=1, 2, \dots, n)$, 将原始方程转化为拉格朗日对偶问题, 如下所示。

$$\max_{\alpha} Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j, \quad (17)$$

$$s. t., \alpha_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \alpha_i y_i = 0, \quad (18)$$

$$\alpha_i (y_i (w^T x_i + b) - 1) = 0. \quad (19)$$

依据 KKT^[16] 最优化条件得到公式(19)所示关系, 则 $\alpha_i > 0$ 时的数据点即为支持向量。利用拉格朗日函数对 w 求偏导可以得到 $w = \sum_{i=1}^n \alpha_i y_i x_i$, 其中 n 表示支持向量个数, 进而依据公式(19)求得参数 b 。由此得到测试样本集的线性判别函数如公式(20)所示。

$$g(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i x_i^T x + b). \quad (20)$$

为了使训练样本实现线性可分, 采用参数较少, 且计算复杂度相对较低的 RBF 核函数作为映射函数 $\Phi(x_i)$, 将原始向量投影到高维空间, 其数学模型如公式(21)所示:

$$K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \exp(-\gamma \|x - x_i\|^2). \quad (21)$$

此外, 为了提高模型的判别性能, 提出了一种自适应变异粒子群算法 (AMPSO) 以实现对其核函数带宽 γ 和惩罚因子 C 的寻优。

3.2 自适应变异 SVM 参数优化

粒子群优化算法是由 Eberhart 与 Kennedy 以鸟群捕食行为为参照提出的一种仿生智能算法, 为了改善算法的收敛性能, 在速度进化方程中引入惯性权重。

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_{1j}(t)(p_{ij}(t) - x_{ij}(t)) + c_2 r_{2j}(t)(p_{gj}(t) - x_{ij}(t)), \quad (22)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (23)$$

其中 i 为粒子标号; j 为维数标号; t 是迭代次数; c_1, c_2 为加速常数; r_{1j} 与 r_{2j} 是两个相互独立的随机函数, 取值为 0-1 之间; p_{ij} 和 p_{gj} 分别代表粒子当前个体最优位置和群体最优位置; ω 表示惯性权重, 为提高 PSO 的全局搜索和局部探索性能, 采用自适应权重的粒子群算法, 其优化公式如下。

$$\omega = \begin{cases} \omega_{\min} + \frac{(\omega_{\max} - \omega_{\min}) \times (f - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f \leq f_{\text{avg}}, \\ \omega_{\max}, & f > f_{\text{avg}} \end{cases}, \quad (24)$$

式中 ω_{\max} 与 ω_{\min} 为权重的最大和最小值, 分别设置为 0.9 和 0.4^[12]; f 为当前粒子的适应度, f_{avg} 与 f_{\min} 分别代表 $t-1$ 代时的平均适应度和最小适应度。对于 $f > f_{\text{avg}}$ 的粒子, 其 ω 取值为 ω_{\max} ; 其它情况下, 在粒子适应度趋于一致时, ω_t 的数值将变大, 反之 ω_t 的数值将变小。

RBF 核的 SVM 参数寻优是一个二维优化问题, 其中惩罚因子 C 为正数, 是用于权衡损失与分类间隔的权重, 其取值过大会增强对于错分样本的惩罚, 虽能够实现对训练集的正确分类, 但将导致过拟合从而影响模型的泛化能力。依据实验经验, 在动作识别过程中, 若 C 的取值大于 100 会出现明显的过拟合现象, 因此根据文献[5]、[17]和[18]设置 C 的上限值为 100。LIBSVM 中默认 C 的取值为 1, 为了提高参数寻优的严谨性, 设置寻优下限为 0.1, 则对于任意的 i 和 j , $C \in [x_{\min}, x_{\max}] = [0.1, 100]$ 。 γ 的数值与支持向量的个数成反比, 能够间接描述数据映射在特征空间的分布, 且支持向量的个数对样本训练和预测的速度具有较大影响, 因而设 $[x_{\gamma\min}, x_{\gamma\max}] = [0.1, 1000]$ 。

依据遗传算法中, 变异算子模拟生物遗传和

进化现象产生新物种的原理,针对种群多样性逐代变化的特点,在自适应权重粒子群算法中引入变异算子,并构建粒子聚集度模型,实现对每代粒子变异概率的动态调节。其可在保证粒子多样性的同时,有效提升算法的全局寻优性能。在变异概率设置方面,利用标准差 σ 描述种群个体之间的离散程度,并对 σ 进行归一化处理,令粒子聚集度为 α ,则第 t 代的 α 计算公式为:

$$\alpha = 1 - \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (f_i^t - f_{avg}^t)^2} - \sigma_{min}}{\sigma_{max} - \sigma_{min}}, \quad (25)$$

式中 f_i^t 为第 t 代时种群中各粒子的适应度, N 为种群规模, f_{avg}^t 代表第 t 代的平均适应度, σ_{max} 与 σ_{min} 分别为全部粒子适应度的标准差最大值和最小值。依据上式所求种群聚集度定义 $t+1$ 代的粒子变异概率公式如下。

$$P_{t+1} = \frac{\lambda \cdot \alpha}{\ln(N \cdot m)}. \quad (26)$$

P_{t+1} 与粒子聚集度 α 成正比, λ 是固定常数的比例增益,取值区间为 $[3, 5]$ 。 m 表示问题定义域的维数,随着问题规模的变大粒子变异的概率将逐渐减小。本文对 $t+1$ 代种群采用粒子概率变异的方式处理,具体执行规则如下:

(1) 变量设定: $rand_n, n \in \{1, 2, 3, 4\}$ 为开区间 $(0, 1)$ 上满足均匀分布的随机数; i 为粒子标号

(2) 基于粒子聚集度 α 与 P_{t+1} 执行变异

(3) if $rand_1 \in (0, P_{t+1}]$

(4) $k = \text{ceil}(2rand_2)$

(5) end if

(6) if $k == 1$

(7) $x_{ik} = 0.5x_{max} \times rand_3 + 1$

(8) end if

(9) if $k == 2$

(10) $x_{ik} = (x_{max} - x_{min}) \times rand_4 + x_{min}$

(11) end if

其中的 $k = \text{ceil}(x)$ 函数可以令 k 等于数轴正方向上与 x 最接近的整数。

3.3 算法具体实现

所提 AMPSO 算法的具体实现步骤为:

Step1: 对粒子群算法的各个参数进行初始化操作,并在定义域内随机选取 N 个粒子的初始位置和速度;

Step2: 依据各粒子初始位置,以 SVM 交叉

验证的分类准确率作为适应度,计算得到初始 p_{ij} 和 p_{gj} ;

Step3: 由公式(26)计算得出当前代的粒子变异概率,并根据(22)~(24)更新种群中各粒子的位置和速度;

Step4: 按照所提规则执行粒子变异;

Step5: 计算粒子适应度,更新 p_{ij} 和 p_{gj} ;

Step6: 若未达到结束条件,则返回 Step3,且代数 $t=t+1$; 否则,输出最优参数 C 和 γ 。

4 实验分析

本部分首先对视频动作数据库进行说明,并对所提 AMPSO 算法在 SVM 参数寻优方面的性能进行分析,通过设计对比实验对所提动作识别方法的有效性进行了验证。

4.1 实验数据库及耗时分析

实验在文献[3]提出的 KTH 数据集和文献[19]提出的 HMDB51 数据集上进行。其中 KTH 数据集包含 2 391 段视频序列,由 25 个实验个体在 4 种背景相对单一的场景下分别执行全部动作而得到,涉及 6 种人体日常行为,即走、慢跑、跑、拳击、挥手和鼓掌。HMDB51 数据集包含 6 766 段视频序列,涉及 51 种人体动作,每种动作至少具有 100 个视频样本。由于该数据集的数据来源主要是电影场景和网络视频,使得类内差别较大,是目前最具挑战性的动作识别数据集。图 2(a)为 KTH 数据集视频帧样本,图 2(b)、(c)为部分 HMDB51 数据集视频帧示意图。

本文的硬件环境为 Intel Core i5-4590 CPU, 3.30 GHz, 4 GB 内存;软件环境为 Ubuntu 14.04 系统下 Matlab2015b。模型训练过程中的耗时主要有以下 2 个部分:(1) STIP 特征提取阶段, KTH 数据集的平均处理速度为 37.224 3 帧/秒, HMDB51 数据集由于镜头晃动、交互动作较多且动作的随意性强,导致产生的特征点较多,平均处理速度为 9.266 5 帧/秒,各视频段的特征提取时间与视频段长度成正比。KTH 数据集视频时长为 15 s 左右, HMDB51 数据集视频时长一般不超过 4 s;(2) 各动作类别 SVM 参数寻优阶段, KTH 数据集下 SVM 单次训练时间平均为 6.23 s,则 5 折交叉验证下适应度 f 耗时平均为 31.15 s。

HMDB51 数据集 SVM 单次训练时间平均为



图 2 KTH 和 HMDB51 数据集样本帧

Fig. 2 Sample frames of KTH and HMDB51 datasets

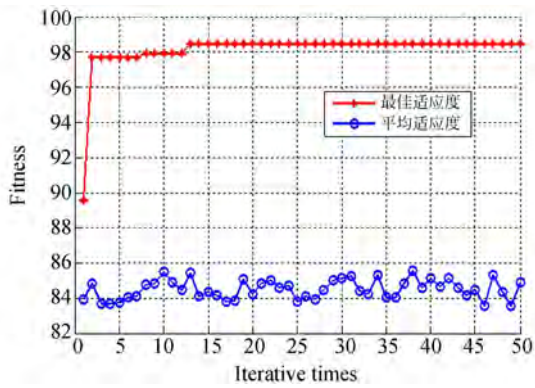
3.03 s, 其 f 平均计算时间 f_T 为 15.16 s。个体动作类别的参数寻优计算复杂度 T 取决于 AMPSO 算法的种群规模 N 和迭代次数 n , 即 $T = f_T \times N \times n$ 。由于对各动作类别的 SVM 参数寻优过程是相互独立的, 因此本文采用多台机器并行计算的策略。在测试分类阶段, 单一模型判别某一视频段动作类别的平均耗时为 6.83 ms, 由于采用一对多策略解决多分类问题, 则分类过程的总体计算复杂度为 $6.83 \text{ ms} \times M$, M 表示动作类别数量。

4.2 SVM 参数寻优分析

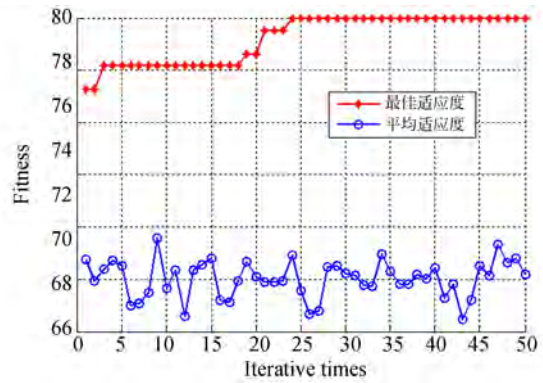
在 KTH 和 HMDB51 数据集上利用 AMPSO 算法对 RBF-SVM 的参数进行寻优, 以验证所提方法的有效性。对于 KTH 数据集, 首先利用 2.1 和 2.2 部分方法分别计算得到全部视频段的中层语义表达, 并将第 2、3、5、6、7、8、9、10 和 22 实验体的视频段 Fisher 向量作为测试样本集, 其余为训练集^[3]。在模型参数 γ 和 C 的寻优过程中, 以训练集中当前动作类别为正样本, 其余为负样本。利用 LIBSVM 工具箱分别对 6 种行为训练相应的二分类模型。设置 $c_1 = 1.5$, $c_2 = 1.7$, 种群规模 $N = 20$, α_{\max} 与 α_{\min} 分别为 50 和 0, λ 设为 3, 采用 5 折交叉验证获取不同参数下的分类准确率作为当前粒子适应度。此外, 由于每个动作视频段的 Fisher 向量为 41 984 维, 且训练集样本数量庞大, 为了减少时间

成本, 在参数寻优过程中设置终止迭代次数为 50。通过 3.3 部分所提算法步骤分别获取 6 个分类模型的最优参数 γ 和 C 。对于 HMDB51 数据集, 利用文献[19]所提供的 train-test 进行算法测试, 为了减小训练规模和保持样本平衡, 在参数寻优过程中设置训练集中当前动作类别为正样本, 从其余 50 类中每类随机抽取 3 个样本组成负样本, 其它参数设置与 KTH 数据集相同。两个数据库的部分参数寻优适应度曲线如图 3 所示。

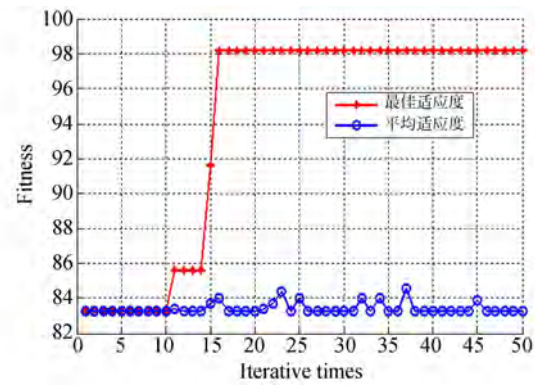
分析图 3 可以得到以下三点结论: (1) 由于 KTH 数据集只包含 6 种动作类别, 类间差别相对明显, 应用训练集样本所得的全局最优适应度保持在 98% 左右; HMDB51 数据集由于动作类别繁多, 存在类内差异大, 类间差别相对较小的问题, 导致各分类模型的最优适应度基本保持在 75%~85% 之间。(2) 从图 3 的适应度曲线变化趋势可以得出, 所提算法在陷入局部最优时能够自适应地提高粒子变异能力, 经过一定次数的迭代即可跳出局部最优, 具备较强的全局寻优性能。(3) 图 3(a)、(b) 的收敛代数分别为 13 和 16; 图 3(c)、(d)、(e)、(f) 收敛代数分别为 19、17、24 和 25。其余动作类别收敛代数均介于 10~35 之间, 从而证明了所提算法能够在迭代次数较少的情况下实现收敛, 为后续的动作分类提供有效支持。



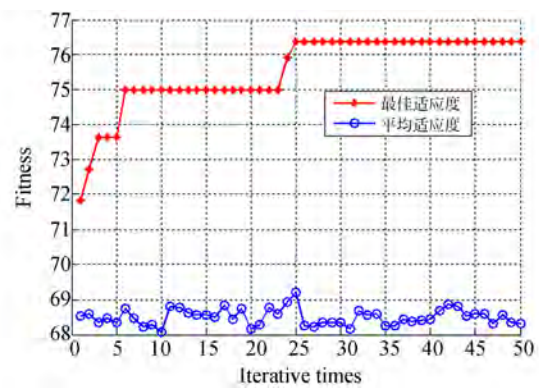
(a) KTH-Boxing $C=10.7702$ $\gamma=3.1146$



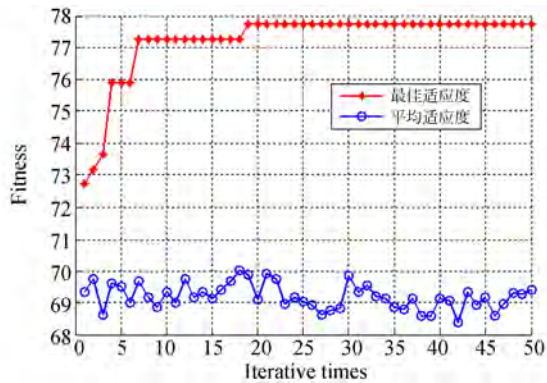
(e) HMDB51-Talk $C=7.0417$ $\gamma=39.7389$



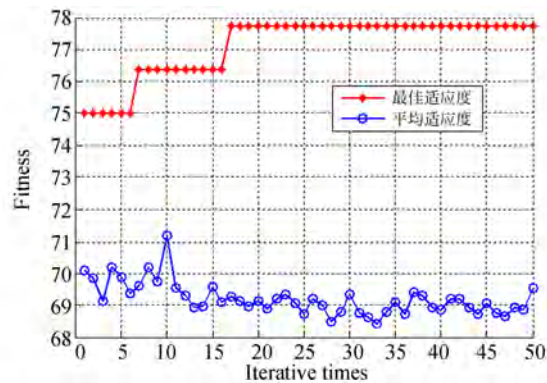
(b) KTH-Waving $C=2.4723$ $\gamma=5.9169$



(f) HMDB51-Wave $C=3.6488$ $\gamma=36.9793$



(c) HMDB51-Chew $C=7.2004$ $\gamma=58.2916$



(d) HMDB51-Kick ball $C=7.4002$ $\gamma=31.7698$

图 3 部分参数寻优适应度曲线

Fig. 3 Partial fitness curves of parameter optimization

整体而言,所提自适应变异粒子群算法 (AMPSO)能够有效避免种群陷入局部最优,具有良好的全局寻优能力和收敛性能。

4.3 识别准确率分析

结合前两节的部分内容,对所提动作识别框架进行对比分析,具体步骤如下:

Step1:训练高斯混合模型。在对两个数据集进行测试时,将高斯混合模型的单元数均设置为 256,分别在 KTH 和 HMDB51 的训练集中随机提取 280 000 和 500 000 个 STIP 特征,经 L1 标准化^[15]处理,并利用 PCA 降维(维数为原先的 50%)后,用于训练得到各自的 GMM。

Step2:获取全部训练集和测试集视频段的中层表达。先对视频特征进行 L1 标准化及 50% PCA 降维,在 GMM 基础上利用 2.2 所述方法得到 41 984 维 Fisher 向量,由公式(13)~(14)得到视频中层表述。

Step3:动作分类。采用一对多方法解决多分类问题,并取预测概率最高的模型所对应的动作

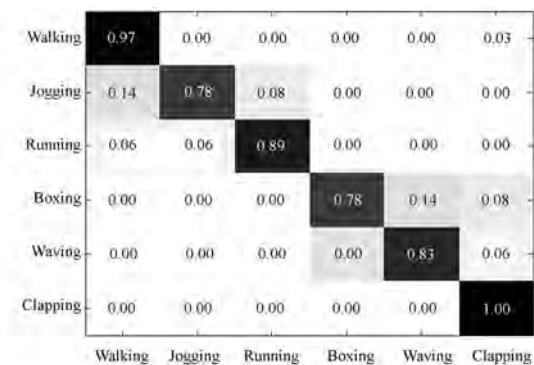


图 4 KTH 数据集动作识别混淆矩阵

Fig. 4 Confusion matrix of action recognition for KTH

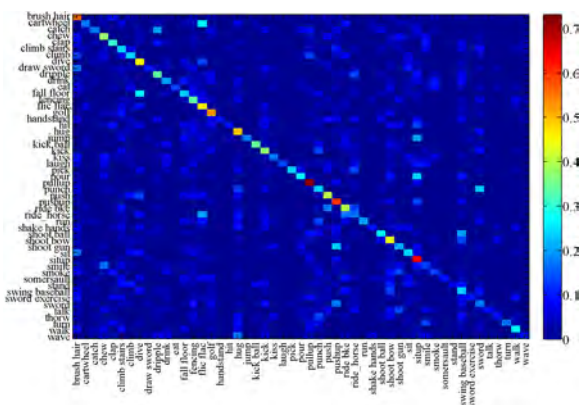


图 5 HMDB51 数据集动作识别混淆矩阵

Fig. 5 Confusion matrix of action recognition for HMDB51

作为当前视频所属类别。其中各动作的 RBF-SVM 模型参数 γ 和 C 分别通过上述 AMPSO 算法优化得到。利用两个数据库的测试集分别对本文所提识别方法进行验证, 所得准确率混淆矩阵如图 4、图 5 所示。

两个数据库所得分类准确率分别为 87.50% 和 26.41%。为了进一步说明所提识别框架的有效性, 采用固定参数下的线性 SVM 和 RBF-SVM 对两个数据库的分类准确率进行测试。对于线性 SVM, 参照文献[5]、[17]和[18]等将参数 C 设置

表 1 与 2 种固定参数下 SVM 识别结果对比

Tab. 1 Comparison with the results based on two fixed SVM (%)

Method type	KTH	HMDB51
Linear SVM	82.87	22.16
RBF-SVM	83.80	20.93
Our method	87.50	26.41

为 100; 对于 RBF-SVM, 训练集采取 5 折交叉验证^[19,20] 获取最优参数 C 和 γ , 得到 KTH 的最优固定参数为 $C=8, \gamma=2$; HMDB51 的最优固定参数为 $C=5, \gamma=68$ 。对比结果如表 1 所示。

由图 4、图 5 和表 1 可知, 分别对各动作的分类模型进行参数寻优, 能够有效增强模型对于所属动作类别的判别性能, 减少动作的类间混淆频率, 进而有效提高了识别方法对于整体数据集的识别准确率。相比于固定参数下的线性 SVM 和 RBF-SVM, 本文方法在 KTH 数据集上所得识别准确率分别提高了 4.63% 和 3.70%; 在 HMDB51 数据集上分别提高了 4.25% 和 5.48%。

由于本文所提识别方法原理上与基于词袋模型的动作识别方法类似, 为了说明本文方法的先进性, 将所提识别框架与较为流行的基于 BOW 的动作识别方法进行比较。动作特征采用相应文献中所使用的人体局部轮廓 (Local Contour, LC)^[1]、HOG、HOF 和 C2^[19] 特征, 对比结果如表 2 所示。

其中, RF 为随机森林 (Random Forest), TM 为模板匹配 (Template Matching)。对比方法中 SVM 均采用 RBF 核函数, 其对应的参数 γ 和 C 通过 5 折交叉验证获取。由表 2 可知, 本文所提算法在 KTH 和 HMDB51 数据库上的测试结果优于上述词袋模型框架下的其它方法。

表 2 与基于 BOW 方法的识别结果对比

Tab. 2 Comparison of recognition results between proposed method and methods based on BOW (%)

Method type	KTH	HMDB51	Ref
LC+BOW+RF-1	82.30	-	[1]
STIP+BOW+TM	81.00	-	[21]
HOG+BOW+SVM	79.69	14.77	-
HOF+BOW+SVM	80.92	17.09	-
STIP+BOW+SVM	-	20.44	[19]
C2+BOW+SVM	-	22.83	[19]
Our method	87.50	26.41	

5 结 论

本文提出一种基于变异粒子群算法的动作识别方法。该方法首先对视频段提取 STIP 特征, 并采用 Fisher 编码获取视频中层语义表达; 针对动作识别问题中一对多框架下固定训练参数导致的 SVM 模型泛化能力较弱的问题, 提出一种变

参数模型训练方法;利用所提 AMPSO 算法对训练过程中每个模型的参数进行寻优,切实提高个体动作分类模型的判别性能。通过实验证明了以下结论:(1) 所提 AMPSO 算法能有效避免种群陷入局部最优,具有较强的全局寻优能力;(2) 在线性 SVM 参数选取时,依据试验经验选取参数 $C=100$ 可以得到较为理想的识别准确率。在利用 RBF-SVM 进行识别时,若随机选取参数 C 和 γ ,则所得准确率往往低于 $C=100$ 时的线性

SVM,但经交叉验证对参数进行寻优之后 2 种方法可以得到较为接近的准确率,差距在 $\pm 2\%$ 以内。(3) 所提识别框架得到的准确率优于 2 种固定参数下的识别方法,在 KTH 和 HMDB51 数据集上分别达到了 87.50% 和 26.41%。

在后续工作中将对时空特征及编码方法进行深入研究,从而获取更具判别力的动作中层表达,进一步提高动作识别性能。

参考文献:

- [1] 王世刚,鲁奉军,赵文婷,等. 应用在线随机森林投票的动作识别[J]. 光学精密工程, 2016, 24(8): 2010-2017.
WANG S G, LU F J, ZHAO W T, *et al.*. Action recognition based on-line random forest voting [J]. *Opt. Precision Eng.*, 2016, 24(08): 2010-2017. (in Chinese)
- [2] PENG X J, WANG L M, WANG X X, *et al.*. Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice [J]. *Computer Vision and Image Understanding*, 2016, 150: 109-125.
- [3] SCHULDT C, LAPTEV I, CAPUTO B. Recognizing human actions: a local SVM approach [C]. *2004 IEEE Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, UK, ICPR, 2004: 32-36.
- [4] LAPTEV I. On space-time interest points [J]. *International Journal of Computer Vision*, 2005, 64(2-3): 107-123.
- [5] WANG H, SCHMID C. Action recognition with improved trajectories [C]. *2013 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Portland, USA, CVPR, 2013: 3551-3558.
- [6] LIU L, WANG L, LIU X. In defense of soft-assignment coding [C]. *2011 International Conference on Computer Vision*, Barcelona, Spain, ICCV, 2011: 2486-2493.
- [7] PERRONNIN F, SANCHEZ J, MENSINK T. Improving the fisher kernel for large-scale image classification [C]. *2010 European conference on computer vision*, Crete, Greece, ECCV, 2010: 143-156.
- [8] 宋悦,林志贤,姚剑敏. 基于 ϵ -支持向量回归机的背光源亮度均匀性评价方法[J]. 液晶与显示, 2015, 30(05): 857-863.
SONG Y, LIN Z X, YAO J M. Luminance uniformity evaluation for backlight based on ϵ -support vector regression [J]. *Chinese Journal of Liquid Crystals and Displays*, 2015, 30(05): 857-863. (in Chinese)
- [9] SHEN L, CHEN H, YU Z, *et al.*. Evolving support vector machines using fruit fly optimization for medical data classification [J]. *Knowledge-Based Systems*, 2016, 96: 61-75.
- [10] HUANG C L, WANG C J. A GA-based feature selection and parameters optimization for support vector machines [J]. *Expert Systems with Applications*, 2006, 31(2): 231-240.
- [11] WANG G P, CHEN S Y, LIU J. A Short-Term prediction model based on support vector regression optimized by artificial Fish-Swarm algorithm [J]. *International Journal of Control and Automation*, 2015, 8(7): 237-250.
- [12] 段海滨,张祥银,徐春芳. 仿生智能计算[M]. 北京: 科学出版社, 2011.
DUAN H B, ZHANG X Y, XU C F. *Bio-inspired Computing* [M]. Beijing: Science Press, 2011. (in Chinese)
- [13] 刘志强,尹建芹,张玲,等. 基于 Kinect 数据主成分分析的人体动作识别[J]. 光学精密工程, 2015, 23(10): 702-711.
LIU Z Q, YIN J Q, ZHANG L, *et al.*. Human action recognition based on Kinect data principal component analysis [J]. *Opt. Precision Eng.*, 2015, 23(10): 702-711. (in Chinese)
- [14] 于海琦,刘真. 一种基于 GA-BP 神经网络结合 PCA 的 LCD 显示器光谱特征化模型[J]. 液晶与显示, 2016, (02): 201-207.

- YU H Q, LIU Z. Spectral characterization model of liquid crystal display based on GA-BP network and PCA [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(02): 201-207. (in Chinese)
- [15] SANCHEZ J, PERRONNIN F, MENSINK T, *et al.*. Image classification with the fisher vector: Theory and practice [J]. *International Journal of Computer Vision*, 2013, 105(3): 222-245.
- [16] ABOUHAWWASH M, SEADA H, DEB K. Towards faster convergence of evolutionary multi-criterion optimization algorithms using karush kuhn tucker optimality based local search [J]. *Computers & Operations Research*, 2017, 79(1): 331-346.
- [17] WANG L, QIAO Y, TANG X. Action recognition with trajectory-pooled deep-convolutional descriptors [C]. 2015 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Boston, USA, CVPR, 2015: 4305-4314.
- [18] LAN Z, LIN M, LI X, *et al.*. Beyond gaussian pyramid: Multi-skip feature stacking for action recognition [C]. 2015 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Boston, USA, CVPR, 2015: 204-212.
- [19] KUEHNE H, JHUANG H, STIEFELHAGEN R. HMDB51: A large video database for human motion recognition [J]. *High Performance Computing in Science and Engineering*, 2013, 12(1): 571-582.
- [20] FERNANDO B, GAVVES S, MOGROVEJO O, *et al.*. Rank pooling for action recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 773-787.
- [21] 王世刚, 孙爱滕. 基于时空兴趣点的单人行为及交互行为识别[J]. *吉林大学学报(工学版)*, 2015, 45(01): 304-308.
- WANG S G, SUN A M. Single and interactive human behavior recognition algorithm based on spatio-temporal interest point [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2015, 45(01): 304-308. (in Chinese)

作者简介:



张国梁(1990—),男,山西临汾人,博士研究生,2015年于太原科技大学获得硕士学位,主要从事机器人分散控制、人体动作识别及智能设备方面的研究。
E-mail: 285719262@qq.com

导师简介:



贾松敏(1964—),女,北京人,博士,教授,2002年于日本国立电气通信大学获得博士学位,主要从事智能服务机器人,计算机视觉等方面的研究。
E-mail: jsm@bjut.edu.cn