基于局部轮廓和随机森林的人体行为识别

蔡加欣[1,2] 冯国灿[1,2] 汤 鑫[1,2] 罗志宏[3]

[1] 中山大学数学与计算科学学院，广东 广州 510275
[2] 广东省计算科学重点实验室，广东 广州 510275
[3] 中山大学信息科学与技术科学学院，广东 广州 510275

摘要 基于视频信息的人体行为识别得到了越来越多的关注。针对人体行为的局部表达，提出了一种新的局部轮廓特征来描述人体的外观姿势，可以同时利用水平和竖直方向上的轮廓变化信息。该特征能有效区分不同动作，与轮廓起始点无关，具有平移、尺度和旋转不变性。针对该特征，提出了一种基于随机森林的两阶段分类方法，使用随机森林分类器对行为视频的局部轮廓进行初分类，并根据每个局部轮廓对应决策类的分类树数目占总分类树数目的比例，提出了一种基于袋外(OOB)数据误差加权投票准则的行为视频分类算法。在测试数据集上的实验结果证实了该方法的有效性。

关键词 机器视觉；行为识别；轮廓特征；随机森林；袋外误差

中图分类号 TP391 **文献标识码** A **doi**：10.3788/AOS201434.1015006

Human Action Recognition Based on Local Image Contour and Random Forest

Cai Jiaxin[1,2] Feng Guocan[1,2] Tang Xin[1,2] Luo Zhihong[3]

[1] *School of Mathematics and Computing Science, Sun Yat-Sen University, Guangzhou, Guangdong* 510275, *China*
[2] *Guangdong Province Key Laboratory of Computational Science, Guangzhou, Guangdong* 510275, *China*
[3] *School of Informational Science and Technology, Sun Yat-Sen University, Guangzhou, Guangdong* 510275, *China*

Abstract Human action recognition in videos has attracted more and more attentions. In view of the local expression of human behavior, a novel local contour feature representing body posture is proposed, which can make full use of information of the contour variation along both horizontal and vertical direction. The proposed local feature can distinguish different actions and is invariant to translation, scaling, rotation and change of start point of human contour. A two stage classifying framework based on random forest is also proposed by using this novel local body contour feature. Random forest is employed to classify each frame of the test video. After that, a video classification method based on out of bag(OOB) error weighted voting strategy to recognize action video according to the ratio of decision trees belonging to each local contour to total decision trees is proposed. Experimental results on test data set prove the effectiveness of proposed method.

Key words machine vision; action recognition; contour feature; random forest; out of bag error

OCIS codes 150.0155; 100.4999; 100.5010; 140.1135

1 引 言

基于视频的人体运动分析最近得到了越来越多的关注，已经成为计算机视觉领域的一个研究热点。特别是对人体行为的识别，在智能监控、增强现实、视频标注、人机交互和体感游戏等方面都具有广泛的应用[1-2]。然而由于光照变化、场景复杂性、动态

收稿日期：2014-03-26；**收到修改稿日期**：2014-06-23

基金项目：国家自然科学基金(61272338)

作者简介：蔡加欣(1988—)，男，博士研究生，主要从事机器视觉和医学图像处理方面的研究。E-mail：caijxin@mail2.sysu.edu.cn

导师简介：冯国灿(1962—)，男，博士，教授，博士生导师，主要从事机器视觉和生物特征识别方面的研究。E-mail：mcsfgc@mail.sysu.edu.cn

遮挡以及人体姿势外观改变等因素，构建一个有效实用的人体行为识别系统仍然是一个挑战。

人体行为的特征表示通常可以分为基于全局的表示和基于局部的表示。在全局表示中，行为的动作特征是在整个视频上进行提取，通常使用一个简单的相似性度量或距离来对视频进行识别。Bobick等[3]提出了运动能力图像和运动历史图像来表示某段时期内的运动事件。Weinland等[4]提出运动历史体积体来表示任意视角的人体动作。Cai等[5]提出一种方向性全局特征来描述人体的往返运动和区分形状相似的人体动作。与局部特征相比，全局特征的优点在于能够以更小的计算代价产生更稳健的特征表达，并且通常只需要用一个简单的基于距离的分类器就可以完成识别任务。局部表示则考虑动态变化，从原始视频数据中提取局部特征，进而通过词袋模型(BOW)、直方图或时间序列等方法来构建视频特征。常用的局部特征提取方法有：提取关键帧；计算光流直方图[6]；提取和跟踪特征角点等局部显著目标[7-8]。金标等[9]提出了一种时空语义来识别人的交互行为。Niebles等[10]在时空兴趣点上建立词袋模型，并用概率主题模型对人体行为进行建模。Klaser等[11]采用三维(3D)梯度直方图作为局部时空体的描述特征，取得了良好的效果。Scovanner等[12]引入视频的3D尺度不变特征变换(SIFT)描述子[13-16]来提取局部运动特征。Liu等[17]结合局部时空体和旋转图来学习近邻图作为人体行为的表示。局部特征的优点在于可以有效地融合空域信息和动态变化信息，但是由于通常需要探测显著点或分割目标区域，局部特征在实际应用中容易受遮挡、光照改变、摄像机移动和视角变化等的影响。

从视频中提取人体运动的特征后，通常可以使用两类方法对人体行为进行识别。1)基于距离和匹配的方法；2)基于状态空间模型的方法。前者常用的测度有序列相关系数、编辑距离和动态时间卷曲距离等。匹配法的优点是计算简单方便，缺点是没有体现行为的动态过程。状态空间模型中常用的有隐马尔科夫模型[17]、对偶隐马尔科夫模型[18]、半马尔科夫模型[19]和条件随机场[20]等。状态空间法的优点是充分利用了时序信息，对行为过程进行直接建模，缺点是参数较多，推广性较差，需要复杂的推断和训练过程。基于形状的人体运动分析近期也得到人们的关注，因为一段人体行为视频可以视为人体剪影的时空变化，而剪影的提取通常要比其他特征的提取要简单。Wang等[21]使用局部保留投影学习人体轮廓的低维流形来进行识别。Cheema等[22]从训练图像集中提取具有判别性的关键轮廓。Chaaraoui等[23]学习多视角关键姿势序列和动态时间归整(DTW)距离进行识别。人体的二值剪影图像通常可以通过使用背景建模法[24]检测人体目标区域来获得，在剪影图像上提取闭合边界就可以得到人体轮廓。在得到人体轮廓后，通常的提取轮廓特征的方法是计算轮廓曲线上每个点到质心的距离作为边界特征[25]。

本文提出了一种新的局部轮廓特征来描述人体的外观姿态。由于人体轮廓在竖直方向上的变化幅值要远小于水平方向，传统的方法无法有效使用其竖直方向上的变化信息。本文提出使用加权距离来构建轮廓特征，使其能够同时利用水平和竖直方向上的轮廓变化信息，并提出了一种起始点对准算法，使得不同轮廓的特征与轮廓起始点无关。本文特征的优点在于：实现简单，可以同时利用水平和竖直方向上的轮廓变化信息，能有效区分不同动作，与轮廓起始点无关，具有平移、尺度和旋转不变性。该轮廓特征提取方法采用与传统的基于距离的边界特征[25]相同的轮廓点采样和尺度规范化方法来获得尺度不变性，但是在计算轮廓曲线上每个点到质心的距离时采用加权距离，来同时获得轮廓变化在水平和竖直方向上的有效表示，克服了传统边界特征只能提取水平方向上轮廓变化信息的缺点，并且还采用了一种起始点对齐方法来对齐相同视频下的各个人体轮廓，与传统的边界特征相比还增加了旋转不变性。与文献[21-23]直接在视频特征上进行识别不同，针对人体行为的局部表达，提出了一种两阶段分类方法，使用随机森林分类器对行为视频的局部轮廓进行初分类，并根据每个局部轮廓对应决策类的分类树数目占总分类树数目的比例，提出了一种基于袋外(OOB)数据误差加权投票准则的行为视频分类算法。实验结果证实了该识别框架的有效性。

2 特征提取

该方法属于基于形状的动作识别，采用从二值图像中提取出的人体姿势特征进行识别。提出了一种新的轮廓特征提取方法。给定一个行为视频上的某一帧图像，使用背景建模法将运动人体的目标区域检测出来[24]，得到一个人体剪影图像，并对剪影图像提取闭合边界，得到一个由边界点序列表示的人体轮廓$\{(x_i,y_i)\}_{i=1}^{n}$。通常提取轮廓特征的方法是

计算轮廓曲线上每个点到质心的距离作为边界特征[25]。由于人体轮廓自身的特点，其轮廓点在 x 轴方向上到质心的距离要远小于 y 轴方向上的距离。如图 1 所示，图 1(a)是一个人体剪影的示例，图 1(b)是其在水平方向上的轮廓曲线，图 1(c)是其在竖直方向上的轮廓曲线，图 1(d)将图 1(b)和图 1(c)中的轮廓曲线在同一个纵轴下表示，横轴的前 100 维对应人体剪影在水平方向上的轮廓曲线，后 100 维对应人体剪影在竖直方向上的轮廓曲线。从图 1 中可见人体剪影在竖直方向上的轮廓变化幅度要远大于水平方向。使用欧氏距离作为轮廓点到质心的长度度量会导致轮廓点的距离曲线中来自 x 轴方向的位置信息过少，得到的人体轮廓表示无法有效体现出在其轮廓点在 x 轴方向上的变化。采用一种加权距离[见(2)式]来解决这个问题，以获得更好的轮廓表示效果。对从二值前景图像中提取出的局部轮廓特征，将其变换到这个加权距离空间中，变换后的轮廓特征具有平移不变性。此外，还提出了一种起始点对齐方法，来对齐相同视频下的各个人体轮廓[见(3)式]，使之具有旋转不变性。然后，将对准后的轮廓特征采样到相同长度并单位化，使之具有尺度不变性。

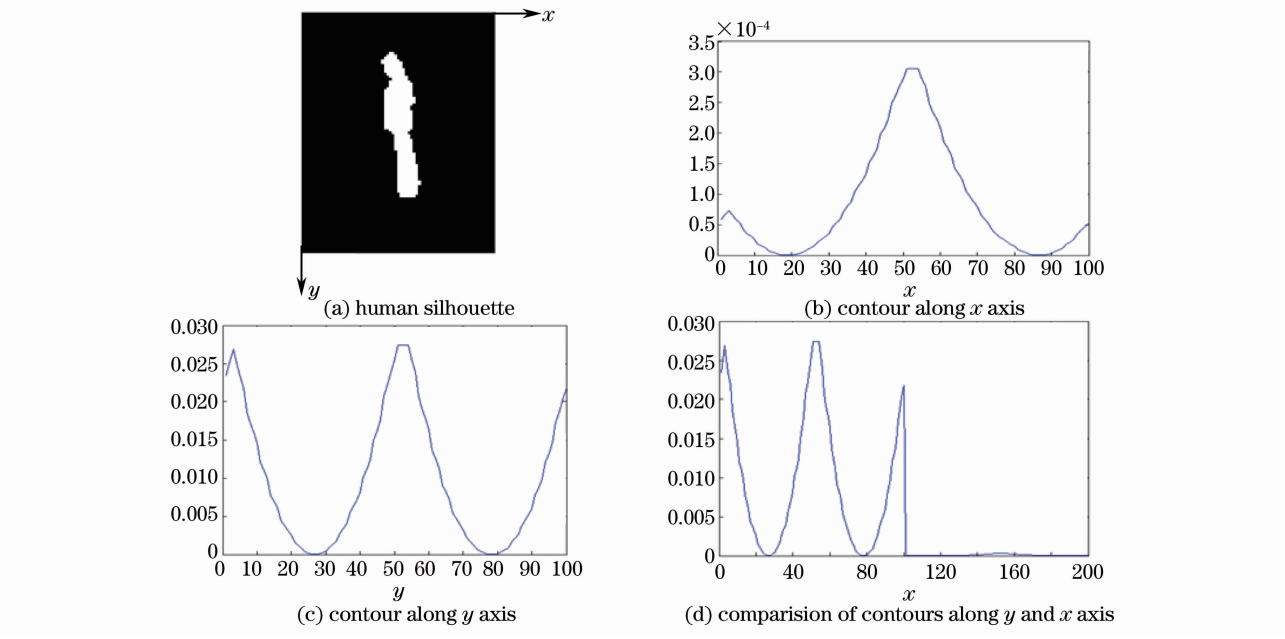

图 1 人体轮廓及轮廓曲线示例

Fig. 1 An example of the body contour and contour curve

具体步骤如下：

1) 给定一个人体轮廓 $\{(x_i,y_i)\}_{i=1}^{n}$，分别计算其在 x 轴和 y 轴方向上轮廓曲线的质心

$$x_c=\frac{\sum_{i=1}^{n}x_i}{n}$$

$$y_c=\frac{\sum_{i=1}^{n}y_i}{n}. \tag{1}$$

2) 计算轮廓曲线上每个点 (x_i,y_i) 到质心 (x_c,y_c) 的加权距离

$$d_i=\sqrt{\frac{(y_i-y_c)^2}{\max_{i=1}^{n}\{(y_i-y_c)^2\}}+\frac{(x_i-x_c)^2}{\max_{i=1}^{n}\{(x_i-x_c)^2\}}}. \tag{2}$$

其中(2)式分别采用水平和竖直两个方向上轮廓点到质心的最大距离来对这两个方向上每个点到质心的距离进行加权。(2)式得到的加权距离与人体轮廓在剪影图像中的位置无关，具有平移不变性。

3) 用(2)式得到的距离向量作为每一帧图像的人体轮廓的初步表示向量。对各个视频下的第 $t+1$ 帧，采用(3)式进行起始点对齐，使之具有旋转不变性，

$$m^*=\arg\min_m \| S_m\boldsymbol{p}_{t+1}-\boldsymbol{p}_t \|^2, \tag{3}$$

式中 $\boldsymbol{p}_t=[d_1^{(t)},\cdots,d_n^{(t)}]$ 表示第 t 帧图像的人体轮廓向量，$\boldsymbol{p}_{t+1}$ 是其相邻帧图像的局部轮廓。S_m 是一个时移算子，$S_m\boldsymbol{p}_t$ 表示将 $\boldsymbol{p}_t$ 时移 m 个单位。

$$S_m\boldsymbol{p}_t=S_m[d_1^{(t)},\cdots,d_n^{(t)}]=[d_{m+1}^{(t)},\cdots,d_n^{(t)},d_1^{(t)},\cdots,d_m^{(t)}]. \tag{4}$$

得到 m^* 后，第 $t+1$ 帧局部轮廓的最佳对齐则为 $S_{m^*}\boldsymbol{p}_{t+1}$。在每个视频下，从 $t=1$ 开始，重复以上方法逐个对每帧图像的人体轮廓进行起始点对齐。因为人体轮廓的旋转在质心对齐后主要表现为轮廓

起始点的时移，所以通过本文方法建立的相同视频下的人体轮廓特征具有旋转不变性。

4）通过采样将对齐后的距离向量 $\boldsymbol{p}=(d_1,d_2,\cdots,d_n)$ 规范到统一的长度 s

$$\tilde{d}_i = d^{\lceil \frac{i*n}{s} \rceil}, \quad i=1,2,\cdots,s. \tag{5}$$

5）将 $\tilde{\boldsymbol{p}}=(\tilde{d}_1,\tilde{d}_2,\cdots,\tilde{d}_s)$ 单位化，使之具有尺度不变性。

$$\hat{d}_i = \frac{\tilde{d}_i}{\sum_{j=1}^{s}\tilde{d}_j}, \quad i=1,2,\cdots,s. \tag{6}$$

将 $\hat{\boldsymbol{p}}=(\hat{d}_1,\hat{d}_2,\cdots,\hat{d}_s)$ 作为人体轮廓的特征表示向量。该特征能有效利用人体轮廓在水平和竖直两个方向上的变化信息，与轮廓起始点无关，具有平移、旋转和尺度不变性。

3　基于随机森林的两阶段识别算法

随机森林（RF）是一种基于多个决策树的集成分类方法，由 Tim 等[26-27]分别独立提出。近年来已经在生物信号处理[28]、人脸识别[29]等领域取得了广泛应用。在随机森林中，先从训练样本中使用有放回的 Bagging 采样（通常是 Boostrap 采样[30]）生成多个训练样本子集，在每个训练样本子集上单独训练一个决策树，未被采样的训练样本 OOB 则用来估计分类器的泛化能力。在每个样本子集上生成对应的决策树后，这些独立同分布的决策树就构成了一个森林，对于一个新的输入样本，森林中的所有决策树被用来判断其归属分类，最终的类别取为这些决策树输出类别的众数。随机森林分类器的优点在于：需要选择的参数较少，实现简单；能够处理高维数据，不需要做特征选择；可以在内部对泛化误差进行无偏估计，具有良好的推广能力；使用采样技术选择样本，可以有效处理类不平衡数据。

与文献[28-29]中直接使用随机森林在已建立的特征上进行分类不同，针对人体行为识别的特点，提出了一种基于随机森林的两阶段分类方法。第一个阶段是对局部轮廓的分类。使用随机森林分类器对行为视频中的每个局部轮廓进行初分类，得到每个局部轮廓的最终决策类别以及将其划分到每一类的决策树数目。第二个阶段是对行为视频的分类，根据每个局部轮廓的输出类别或者其对应每个类的分类树数目，提出了两种投票准则对视频进行分类。

一个行为视频是由若干帧的静态图像组成。我们使用随机森林对所有帧进行分类，采用局部轮廓作为特征，并使用主成分分析将局部轮廓特征降到一定维数 D。训练集取为训练视频的所有帧，每一帧的类别取为其所属视频的归属类别。测试集则是测试视频包含的帧，每一帧的输出类别取为随机森林中决策树输出类别的众数。采用随机森林对图像帧初分类的具体步骤如下：

1）对训练图像集，用 Boostrap 采样生成 N 个子样本集；

2）对每个子训练集，随机选择 m 个属性（$1\leqslant m\leqslant D$）作为节点分裂的候选属性；

3）计算每个子样本集在每个候选节点上的 Gini 指数。其中节点 n 的 Gini 指数的定义为

$$G(n)=\sum_{i\neq j}P(w_i)P(w_j)=1-\sum_{j}P^2(w_j), \tag{7}$$

式中 $P(w_i)$ 是节点 t 处属于第 i 类样本个数占训练样本总数的频度；

4）在每个子样本集上生成一个决策树。对 Gini 指数最大的候选属性进行分裂，然后重新计算 Gini 指数。重复分裂步骤直到 Gini 指数小于预定阈值；

5）训练完所有决策树后，使用第 k 个决策树对未被采样的袋外训练样本进行分类，将得到的分类结果与袋外训练样本的类别相比较，这样第 k 个决策树的 OOB 误差 o_k（$k=1,2,\cdots,N$）就可以通过计算该决策树在袋外数据上的分类误差来得到；

6）对测试视频的每一图像帧用已生成的决策树分类，得到每个决策树对每一帧的分类结果。使用随机森林中决策树输出类别的众数作为每一帧的输出类别。

在得到所有帧的类别后，每个行为视频的类别就可以由之前随机森林算法所得到的帧分类结果来投票决定。提出两种投票策略，分别记为 RF-0 和 RF-1。

1）RF-0

在这种投票方案中，视频的分类结果由每一帧的分类结果投票决定。最终的决策结果是所有帧归属类别的众数。投票过程可记为

$$Y=\arg\max_{j}\sum_{i=1}^{T}h_j(I_i), \tag{8}$$

式中 Y 是视频的分类类标，$h_j(I_i)$ 是一个布尔函数，指示视频的第 i 帧是否被划分到第 j 类。T 是该视频的帧数。

2）RF-1

使用袋外数据误差加权投票，分类结果由视频

中所有帧对应于每一类的决策树数目及决策树的袋外数据误差决定。投票过程如下：

$$Y = \arg\max_j \sum_{k=1}^{N}\sum_{i=1}^{T} w_k \boldsymbol{I}_k[t_j(I_i)=1], \quad (9)$$

式中 $I_k[t_j(I_i)=1]$ 是一个指示函数，如果在随机森林分类过程中第 k 个决策树将视频第 i 帧划分到第 j 类则为 1，否则为 0。权重 w_k 由第 k 个决策树的 OOB 误差 o_k 决定：

$$w_k = \frac{(1-o_k)^2}{\sum_{k=1,N}(1-o_k)^2}. \quad (10)$$

比较这里提出的两种投票准则。RF-0 是基于帧投票，采用视频内每帧的分类结果投票决定视频的分类结果。RF-1 则是基于袋外误差加权的分类树投票，采用视频内所有帧对应于每一类的决策树来对视频的分类结果进行投票，其中每棵决策树的投票值由其袋外数据误差决定。在这里，袋外数据误差的作用是用来衡量各决策树分类结果的信度。通过(10)式，可以对每颗决策树设置不同的权重，使得袋外误差小的决策树具有较大的权重，从而使得在投票时分类置信度越大的决策树对投票结果具有越大的影响。通过袋外误差进行加权投票，能使分类器具有更好的对抗噪声的能力。

4 实验结果与分析

4.1 比较算法

将下面所描述动作识别方法作为本文方法的比较算法。对人体动作视频数据集中某个视频，在其第 t 帧上按(6)式建立局部轮廓特征 $[\hat{d}_1(t),\cdots,\hat{d}_s(t)]$。若该视频的总帧数为 T，则其描述特征取为

$$[\hat{d}_1(1),\cdots,\hat{d}_s(1),\cdots,\hat{d}_1(t),\cdots,\hat{d}_s(t),\cdots,\hat{d}_1(T),\cdots,\hat{d}_s(T)]. \quad (11)$$

按照这种方法对数据集的所有视频建立描述特征。由于不同视频的帧数未规范化到一致，每个视频的描述特征维数都不同。取所有视频的最大描述特征维数为统一的特征向量长度，将各个视频的描述特征向量补零对齐到该长度。然后在视频特征上分别采用随机森林和支持向量机训练模型并对每个测试视频进行识别。为了方便描述，将上面所述的在视频特征上用随机森林和支持向量机(SVM)的识别方法分别记为 RF 和 SVM。

此外，还与采用词袋表示 BOW 作为视频特征的方法进行比较。词袋表示通过以下步骤建立：对(6)式所得到的所有局部轮廓，使用 K 均值聚类算法将其聚为 K 个类；对每个视频，建立一个长度为 K 的特征表示向量，其中第 k 维的值取为该视频中被划分到第 k 个聚类的局部轮廓的数目。将在词袋表示的视频特征上用随机森林和支持向量机进行识别的方法分别记为 BOW+RF 和 BOW+SVM。

对应的，将使用本文提出的局部轮廓特征和采用本文提出的(9)式和(10)式分别进行两阶段随机森林算法进行行为识别的方法称为 RF-0，RF-1。

4.2 Weizmann 数据库

采用 Weizmann 数据库[31]作为实验数据集，来验证该算法的有效性。Weizmann 数据集一共包含 10 个人体动作，分别为 bend，jack，jump，pjump，run，side，skip，walk，wave1 和 wave2。每种动作由 9 个人完成。数据集中一共包含 93 个视频，每个视频的背景和视角均不变，每帧图像的分辨率为 144 pixel×180 pixel，帧速率为 25 fps。图 2 给出了 Weizmann 数据库的部分示例图像及其对应的二值前景图像。对前景图像提取局部轮廓特征，其中轮廓特征向量的规范化长度 s 取为 100，然后用主成分分析法(PCA)将其降到一定维数，默认设置降维后维数为 60。随机森林算法使用的决策树数目 N 和决策树所使用的变量个数 m 分别取为 500 和 6。由于样本数量较少，使用留一法交叉验证来统计识别率。得到 RF-1 的识别率为 91.40%，RF-0 的识别率为 89.25%，分类结果的混淆矩阵分别如图 2、3 所示。另外还测试了相同设置下使用文献[25]特征结合 RF-1 的识别结果，得到识别率为 87.10%，证实了该特征的有效性。将该方法的测试结果与其它方法进行比较，比较结果如表 1 所示。其中文献

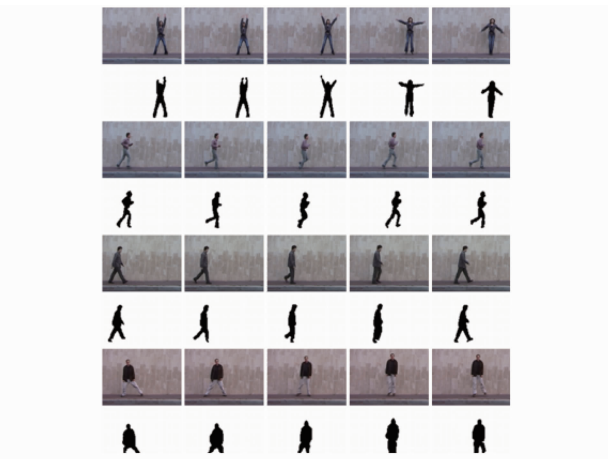

图 2 Weizmann 数据集中部分示例图像及其剪影提取[31]

Fig. 2 Example images and their extracted silhouettes of Weizmann dataset[31]

[10－12,23,32]属于基于局部特征的人体行为识别方法,文献[5]属于基于全局特征的识别方法。文献[5]采用将数据集划分为固定训练集和测试集的方法进行实验,文献[10－12,23,32]采用留一法交叉验证。BOW＋RF 和 BOW＋SVM 方法中,K 取值为 100。实验结果表明本文方法优于所列的其他方法,证明了本方法的有效性。

	bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2
bend	1	0	0	0	0	0	0	0	0	0
jack	0	1	0	0	0	0	0	0	0	0
jump	0.11	0	0.89	0	0	0	0	0	0	0
pjump	0	0.11	0	0.89	0	0	0	0	0	0
run	0	0	0	0	0.90	0	0	0.10	0	0
side	0	0	0	0.11	0	0.89	0	0	0	0
skip	0	0	0	0	0	0	1	0	0	0
walk	0	0	0	0	0	0	0	1	0	0
wave1	0	0.11	0	0	0	0	0	0	0.78	0.11
wave2	0	0.11	0	0	0	0	0	0	0	0.89

图 3　RF-1 在 Weizmann 数据库上分类结果的混淆矩阵

Fig. 3　Confusion matrix of RF-1 on Weizmann dataset

表 1　各方法在 Weizmann 库上实验结果比较

Table 1　Comparison of different methods on Weizmann dataset

Method	Rewgnition rate /%
RF-1	91.40
RF-0	89.25
Feature of Ref.[25]＋RF1	87.10
RF	17.20
SVM	15.05
BOW＋RF	87.10
BOW＋SVM	53.76
Ref.[5]	90.00
Ref.[10]	90.00
Ref.[11]	84.30
Ref.[12]	82.64
Ref.[23]	90.32
Ref.[32]	87.50

另外还比较了各对比算法在分别使用本文特征和文献[25]特征时在 Weizmann 数据库的识别效果。实验结果如图 5 所示。实验结果显示,对大部分识别算法,该局部轮廓特征要优于文献[25]特征。

此外还测试了 RF-1 在 Weizmann 数据库上识别率与 PCA 降维维数的关系,如图 6 所示。此外还测试了该方法的识别率与随机森林算法的两个参数(决策树数目 N 和决策树所使用的变量个数 m)的

	bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2
bend	1	0	0	0	0	0	0	0	0	0
jack	0	1	0	0	0	0	0	0	0	0
jump	0.11	0	0.78	0.11	0	0	0	0	0	0
pjump	0	0.22	0	0.78	0	0	0	0	0	0
run	0	0	0	0	0.90	0	0	0.10	0	0
side	0	0	0	0.11	0	0.89	0	0	0	0
skip	0	0	0	0	0	0	1	0	0	0
walk	0	0	0	0	0	0	0	1	0	0
wave1	0.11	0.11	0	0	0	0	0	0	0.67	0.11
wave2	0	0.11	0	0	0	0	0	0	0	0.89

图 4　RF-0 在 Weizmann 数据库上分类结果的混淆矩阵

Fig. 4　Confusion matrix of RF-0 on Weizmann dataset

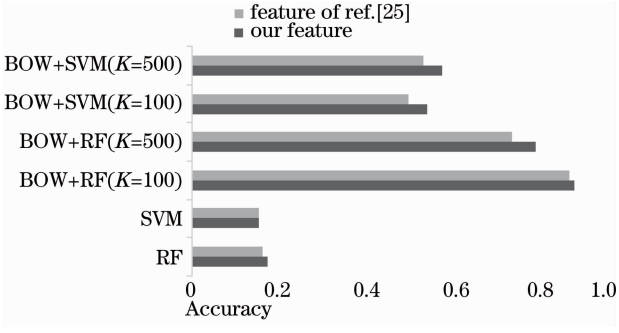

图 5　本文特征和文献[25]特征在 Weizmann 数据库上识别率的比较

Fig. 5　Comparison of our feature with feature of Ref.[25] on Weizmann dataset

关系,如图 7、8 所示。结果显示 RF-1 在大部分情况下识别效果好于 RF-0。实验结果表明该方法对参数具有稳健性,容易寻找有效的参数,简单实用。

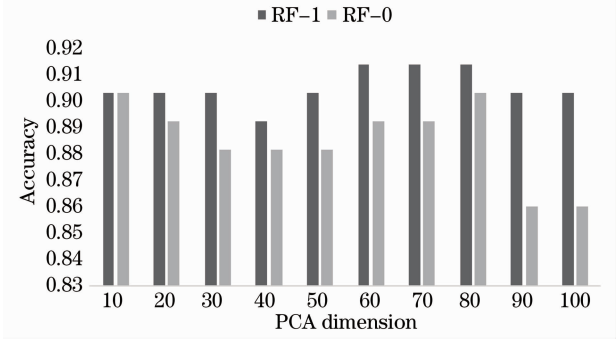

图 6　RF-1 和 RF-0 在 Weizmann 数据库上识别率与 PCA 降维维数的关系

Fig. 6　Relations between accuracy of RF-1, RF-0 and PCA dimension on Weizmann dataset

4.3　MuHAVi-MAS14 数据库

选择多视角的 MuHAVi-MAS14 数据库[33]作为另一个测试数据集,来验证该方法的有效性。

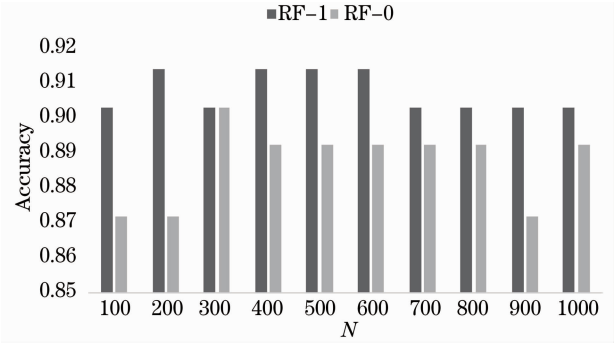

图 7 RF-1 和 RF-0 在 Weizmann 数据库上识别率与随机森林算法分类树数目 N 的关系

Fig. 7 Relation between accuracy of RF-1, RF-0 and tress number N of random forest on Weizmann dataset

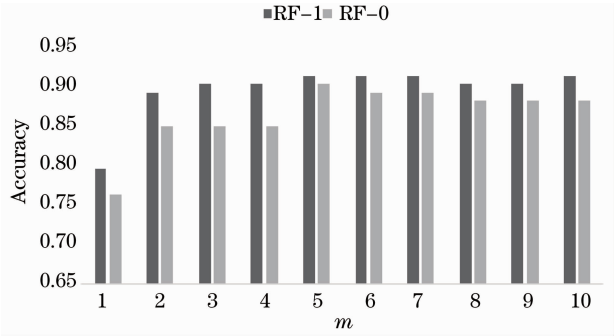

图 8 RF-1 和 RF-0 在 Weizmann 数据库上识别率与随机森林算法决策树变量个数 m 的关系

Fig. 8 Relation between accuracy of RF-1, RF-0 and number of variables m randomly sampled as candidates at each split of random forest on Weizmann dataset

MuHAVi-MAS14 数据集一共含有 14 个动作(Collapse Left, Collapse Right, Guard To Kick, Guard To Punch, Kick Right, Punch Right, Run Left To Right, Run Right To Left, Stand up Left, Stand up Right, Turn Back Left, Turn Back Right, Walk Left To Right 和 Walk Right To Left 等),由 2 个人完成,包含 2 个视角(正面和 45°),每个视角有 68 个视频,共计 136 个视频。MuHAVi-MAS14 数据库中包含了视角变化,所定义的动作间具有较大的混淆性(如 Run Right To Left 和 Run Left To Right 都可以视为 Run),识别难度比 Weizmann 数据库更大。图 9 给出了 MuHAVi-MAS14 数据集的部分示例图像。取轮廓特征向量的规范化长度 s 为 100,主成分分析法(PCA)降维后维数为 20,决策树数目 N 为 500,决策树所使用的变量个数 m 为 7,使用留一法交叉验证来统计识别率。将该方法的测试结果与其他方法进行比较,比较结果如表 2 所示。其中 RF-1 的识别率为 86.03%,RF-0 的识别率为 84.56%,分类结果的混淆矩阵分别如图 10、11 所示。BOW+RF 和 BOW+SVM 方法中,K 取值为 100。文献[25]特征结合 RF-1 方法的识别率为 79.41%。实验结果显示该方法的识别结果优于该库提供的参考识别率 82.35%,证实了该方法的有效性。

图 9 MuHAVi-MAS14 数据集部分示例图像[33]

Fig. 9 Example images of MuHAVi-MAS14 dataset[33]

表 2 各方法在 MuHAVi-MAS14 库上实验结果比较

Table 2 Comparison of different methods on MuHAVi-MAS14 dataset

Method	Recagnition rate /%
RF-1	86.03
RF-0	84.56
Feature of Ref.[25]+RF1	79.41
RF	63.97
SVM	56.62
BOW+RF	74.26
BOW+SVM	69.12
Ref.[32]	82.35

	CL	CR	CK	GP	KR	PR	RLR	RRL	SL	SR	TL	TR	WLR	WRL
CL	1													
CR	0.13	0.75			0.13									
CK			0.69		0.25		0.06							
GP			0.19	0.69		0.06	0.06							
KR					1									
PR						1								
RLR							1							
RRL								1						
SL									0.25	0.75				
SR										1				
TL						0.25	0.25				0.5			
TR												0.75	0.13	0.13
WLR													1	
WRL														1

图 10 RF-1 在 MuHAVi-MAS14 数据库上分类结果的混淆矩阵

Fig. 10 Confusion matrix of RF-1 on MuHAVi-MAS14 dataset

	CL	CR	CK	GP	KR	PR	RLR	RRL	SL	SR	TL	TR	WLR	WRL
CL	1													
CR		0.88			0.13									
CK			0.69	0.06	0.06		0.06	0.06					0.06	
GP			0.19	0.63			0.06			0.13				
KR					1									
PR						1								
RLR							1							
RRL								1						
SL									0.25	0.75				
SR										1				
TL						0.25					0.5			0.25
TR	0.13											0.63	0.13	0.13
WLR													1	
WRL														1

图 11　RF-0 在 MuHAVi-MAS14 数据库上分类结果的混淆矩阵

Fig. 11　Confusion matrix of RF-0 on MuHAVi-MAS14 dataset

另外还比较了各对比算法在分别使用本文特征和文献[25]特征时在 MuHAVi-MAS14 数据库的识别效果。实验结果如图 12 所示。实验结果显示，对大部分识别算法，本文的局部轮廓特征要优于文献[25]特征。

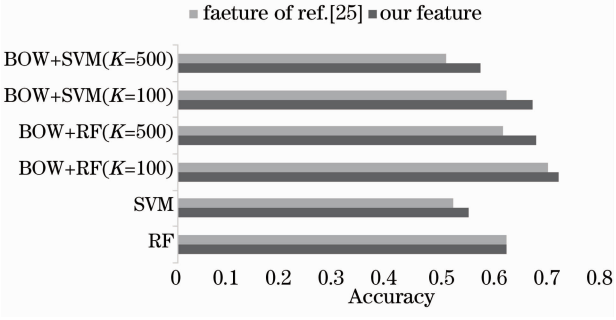

图 12　本文特征和文献[25]特征在 MuHAVi-MAS14 数据库上识别率的比较

Fig. 12　Comparison of our feature with feature of Ref. [25] on MuHAVi-MAS14 dataset

图 13 给出了 MuHAVi-MAS14 数据库上，RF-1 在 m 取为 10 以及决策树数目 N 为 500 时的识别率与 PCA 降维维数的关系。此外还测试了在该方法在 PCA 降维维数为 20 以及决策树数目 N 为 500 时的识别率与决策树所使用的变量个数 m 的关系，如图 14 所示。结果显示本文方法的识别结果稳定，对参数变化不敏感。

4.4　计算耗时分析

在 Intel Core I7-4700MQ 2.40 GHz，4 GB 内存的硬件环境和 Window 8.1，MATLAB8.1 的软件环境下，对本文方法和对比方法在 Weizmann 数据库上的 CPU 运行时间进行分析。RF-0 和 RF-1 在提取局部轮廓特征阶段的平均耗时为 35.81 s，而提取文献[25]特征的平均耗时为 34.80 s。对比方法 RF 和 BOW-RF 需要在提取局部轮廓特征后进一步提取视频特征，该阶段的平均耗时分别为 0.70 s 和 22.90 s。因此 RF 和 BOW-RF 在特征提取的平均耗时分别为 35.50 s 和 57.70 s。SVM 和 BOW-SVM 的特征提取方法分别与 RF 和 BOW-RF 相同。对于方法 RF-0 和 RF-1，在第一阶段生成决策树[步骤 1)～4)]的平均耗时为 10.29 s，计算各决策树的 OOB 误差[步骤 5)]平均耗时 39.35 s，进行帧分类[步骤 6)]的平均耗时为 5.90 ms。RF-0 和 RF-1 在第二阶段的平均耗时分别 0.18 ms 和 7.35 ms。注意到步骤 5)对 RF-0 不是必须的，所以 RF-0 的训练耗时和测试耗时分别为 10.29 s 和 6.08 ms，RF-1 的训练耗时和测试耗时分别为 49.64 s 和 13.25 ms。与之相比，RF 的训练耗时和测试耗时分别为 0.34 和 11.75 ms，SVM 的训练耗时和测试耗时分别为 1.16 s 和 190.70 ms，BOW+

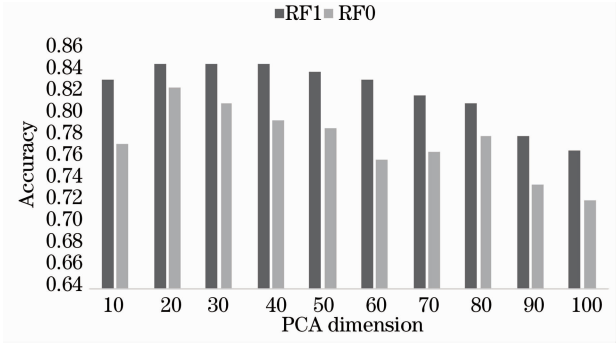

图 13　RF-1 和 RF-0 在 MuHAVi-MAS14 数据库上识别率与 PCA 降维维数的关系

Fig. 13　Relation between accuracy of RF-1, RF-0 and PCA dimension on MuHAVi-MAS14dataset

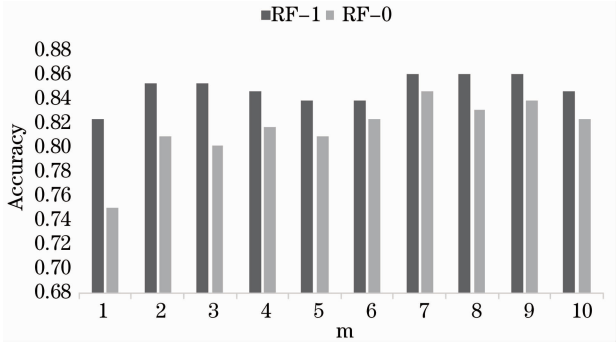

图 14　RF-1 和 RF-0 在 MuHAVi-MAS14 数据库上识别率与随机森林算法决策树变量个数 m 的关系

Fig. 14　Relation between accuracy of RF-1, RF-0 and number of variables m randomly sampled as candidates at each split of random forest on MuHAVi-MAS14 dataset

RF 的训练耗时和测试耗时分别为 0.30 s 和 22.14 ms，BOW+SVM 的训练耗时和测试耗时分别为 0.26 s 和 32.39 ms。图 15 给出了各方法在 Weizmann 数据库上的平均计算耗时分析，从图中可以看出，RF-0 和 RF-1 在特征提取阶段和测试阶段的耗时都明显低于对比方法，只是在训练时间上有所增加。

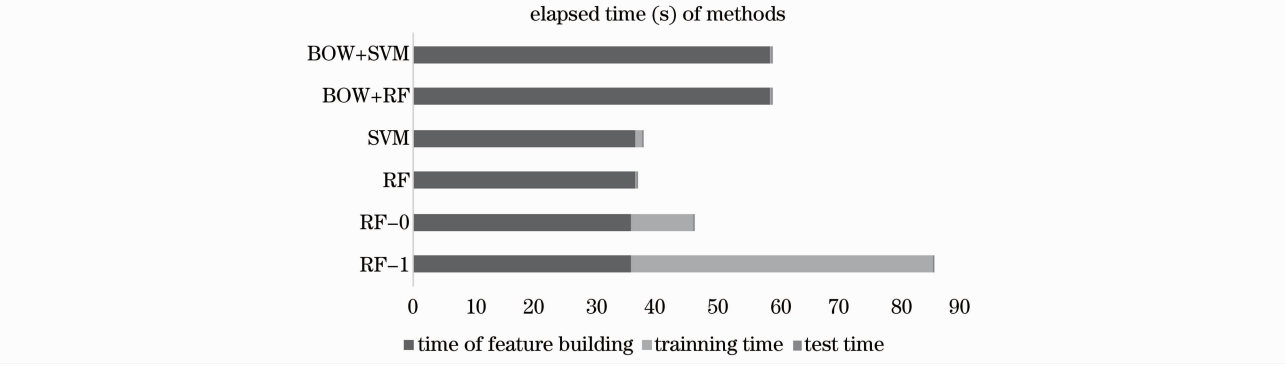

图 15 各方法在 Weizmann 数据库上的平均计算耗时分析

Fig. 15 Average elapsed time of methods on Weizmann dataset

5 结 论

提出了一种基于局部轮廓特征和两阶段随机森林分类的人体动作识别方法。该方法对人体轮廓构建基于加权距离的局部不变特征，并使用随机森林分类器对人体轮廓进行初分类，根据局部轮廓的分类结果使用加权投票准则对人体行为视频进行分类。在测试数据库上的实验结果证实了此方法的有效性。

参 考 文 献

1 Li Hongsong, Li Da. Some advances in human motion analysis [J]. PR&AI, 2009, 22(1): 70－78.
黎洪松，李 达. 人体运动分析研究的若干新进展[J]. 模式识别与人工智能，2009，22(1)：70－78.

2 Xu Guangyou, Cao Yuanyuan. Action recognition and activity understanding: a review [J]. Journal of Image and Graphics, 2009, 14(2): 189－195.
徐光祐，曹媛媛. 动作识别与行为理解综述[J]. 中国图象图形学报，2009，14(2)：189－195.

3 A F Bobick, J W Davis. The recognition of human movement using temporal templates [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(3): 257－267.

4 D Weinland, R Ronfard, E Boyer. Free viewpoint action recognition using motion history volumes [J]. Computer Vision and Image Understanding, 2006, 104(2-3): 249－257.

5 J X Cai, G C Feng, X Tang. Human action recognition using oriented holistic feature [C]. 20th IEEE International Conference on Image Processing, 2013. 2420－2424.

6 Guan Zhiqiang, Chen Qian, Gu Guohua, *et al*.. Dim target detection based on optical flow histgram in low frame frequence in clouds background [J]. Acta Optica Sinica, 2008, 28(8): 1496－1501.
管志强，陈 钱，顾国华，等. 基于光流直方图的云背景下低帧频小目标探测方法[J]. 光学学报，2008，28(8)：1496－1501.

7 Wang Xiangjun, Wang Yan, Li Zhi. Fast target recognition and tracking method based on characteristic corner [J]. Acta Optica Sinica, 2007, 27(2): 360－364.
王向军，王 研，李 智. 基于特征角点的目标跟踪和快速识别算法研究[J]. 光学学报，2007，27(2)：360－364.

8 Gao Lin, Tang Peng, Sheng Peng, *et al*.. Visual object tracking based on conditional random field under complex scene [J]. Acta Optica Sinica, 2010, 30(6): 1721－1728.
高 琳，唐 鹏，盛 鹏，等. 复杂场景下基于条件随机场的视觉目标跟踪[J]. 光学学报，2010，30(6)：1721－1728。

9 Jin Biao, Hu Wenlong, Wang Hongqi. Moving-objects interaction recognition based on the spatial-temporal semantic information [J]. Acta Optica Sinica, 2012, 32(5): 0515002.
金 标，胡文龙，王宏琦. 基于时空语义信息的视频运动目标交互行为识别方法[J]. 光学学报，2012，32(5)：0515002.

10 J C Niebles, H C Wang, F F Li. Unsupervised learning of human action categories using spatial-temporal words [J]. Int J Comput Vis, 2008, (79): 299－318.

11 K Alexander, M Marcin, S Cordelia. A spatio-temporal descriptor based on 3d-gradients [C]. British Machine Vision Conference, IEEE Computer Society, 2008. 995－1004.

12 S Poul, A Saad, S Mubarak. A 3-dimensional sift descriptor and its application to action recognition [C]. Proceedings of the 15th international conference on Multimedia, IEEE Computer Society, 2007. 357－360.

13 Haihua Cui, Wenhe Liao, Ning Dai, *et al*.. Registration and integration algorithm in structured light three-dimensional scanning based on scale-invariant feature matching of multi-source images [J]. Chin Opt Lett, 2012, 10(9): 091001.

14 Jianfang Dou, Jianxun Li. Robust image matching based on SIFT and delaunay triangulation [J]. Chin Opt Lett, 2012, 10(s1): s11001.

15 Mingliang Gao, Xiaomin Yang, Yanmei Yu, *et al*.. Photometric invariant feature descriptor based on SIFT [J]. Chin Opt Lett, 2012, 10(s1): s11003.

16 J G Liu, S Ali, M Shah. Recognizing human actions using multiple features [C]. Computer Vision and Pattern Recognition, IEEE Computer Society, 2008. 1－8.

17 J Yamato, J Ohya, K Ishii. Recognizing human action in time sequential images using hidden Markov model [C]. IEEE Conference on Computer Vis Pattern Recognit, 1992. 379－385.

18 M Brand, N Oliver, A Pentland. Coupled hidden Markov models for complex action recognition [C]. IEEE Conference on

Computer Vis Pattern Recognit, 1997: 994－999.

19 Q F Shi, L Wang, L Cheng, *et al*.. Discriminative human action segmentation and recognition using semi-markov model [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1－5.

20 C Sminchisescu, A Kanaujia, D Metaxax. Conditional models for contextual human motion recognition [J]. Computer Vision and Jmago Understanding, 2006, 104(2-3): 210－220.

21 L Wang, D Suter. Visual learning and recognition of sequential data manifolds with applications to human movement analysis [J]. Computer Vision and Image Understanding, 2008, 110(2): 153－172.

22 S Cheema, A Eweiwi, C Thurau, *et al*.. Action recognition by learning discriminative key poses [C]. IEEE International Conference on Computer Vision Workshops, 2011. 1302－1309.

23 C A Andre, C P Pau, Flórez-Revuelta Francisco. Silhouette-based human action recognition using sequences of key poses [J]. Pattern Recognition Letters, 2013, 34(15): 1799－1807.

24 Ming Ying, Jiang Jingjue. Background modeling and moving-objects detection based on cauchy distribution for video sequence [J]. Acta Optica Sinica, 2008, 28(3): 587－592.
明　英, 蒋晶珏. 基于柯西分布的视频图像序列背景建模和运动目标检测[J]. 光学学报, 2008, 28(3): 587－592.

25 S Suzuki, K Be. Topological structural analysis of digitized binary images by border following[J]. Comput Vision Graphics Image Process, 1985, 30(1): 32－46.

26 H T Kam. The random subspace method for constructing decision forests [J]. Pattern Analysis and Machine, 1998, 20 (8): 832－844.

27 B Leo. Random Forests [J]. Machine Learning, 2001, 45(1): 5－32.

28 Zhang Hongqiang, Liu Guangyuan, Lai Xiangwei. Application of random forest algorithm in important feature selection from EMG signal [J]. Computer Science, 2013, 40(1): 200－202.
张洪强, 刘光远, 赖祥伟. 随机森林算法在肌电的重要特征选择中的应用[J]. 计算机科学, 2013, 40(1): 200－202.

29 Guo Jinxin, Chen Wei. Face recognition based on HOG multi-feature fusion and random forest [J]. Computer Science, 2013, 40(10): 279－282.
郭金鑫, 陈　玮. 基于 HOG 多特征融合与随机森林的人脸识别[J]. 计算机科学, 2013, 40(10): 279－282.

30 Haixia Xu, Xianbin Wen, Yongliao Zou, *et al*.. Performance evaluation and segmentation for synthetic aperture radar image using mixture multiscale autoregressive model and bootstrap technique [J]. Chin Opt Lett, 2012, 10(s1): s11005.

31 B Mosh, G Lena, S Eli, *et al*.. Actions as space-time Shapes [C]. IEEE International Conference on Computer Vision, 2005, 2: 1395－1402.

32 Q Zhao, H S Horace. Unsupervised approximate-semantic vocabulary learning for human action and video classification [J]. Pattern Recognition Letters, 2013, 34(15): 1870－1878.

33 S Singh, S Velatin, H Ragheb. Muhavi: a multicamera human action video dataset for the evalution of action recognition methods [C]. IEEE International Conference on Boston: Adavanced Video and Signal Based Surveillance (AVSS), 2010. 48－55.

栏目编辑: 张浩佳