基于Fisher鉴别字典学习的人体行为识别

季 冲 王 胜 陆建峰
(南京理工大学计算机科学与工程学院 南京 210094)

摘 要 人体行为识别是计算机视觉中的一个重要研究领域,具有广阔的应用前景。研究了基于Fisher鉴别的字典学习方法在人体行为识别上的应用。首先对人体行为的视频序列提取了局部时空特征,并通过随机投影法降维;然后把降维后的特征作为待分类的信号进行Fisher鉴别字典学习,从而增强字典和编码系数的鉴别能力;最后同时利用重构误差和稀疏表示系数进行分类。实验结果验证了所提方法在人体行为识别上的有效性与鲁棒性。

关键词 稀疏表示,人体行为识别,运动特征,Fisher鉴别准则
中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.07.048

Human Action Recognition Based on Fisher Discrimination Dictionary Learning

JI Chong WANG Sheng LU Jian-feng
(Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract Human action recognition is a hot computer vision research field, and has broad application prospects. This paper explored the application of Fisher discrimination based dictionary learning on human action recognition. First, local spatial-temporal features are extracted from the video sequences and random projection is used to reduce dimension. Then, Fisher discrimination based dictionary learning is performed on the reduced motion features. Last, a new classification scheme is proposed using both reconstruction error and representation coefficients. Experimental results confirm the efficiency and robustness of the proposed scheme.

Keywords Sparse representation, Human action recognition, Motion features, Fisher discrimination criterion

1 引言

人体行为识别的主要任务是利用计算机视觉技术对视频序列中处于运动状态的人进行检测、跟踪,并识别其行为。人体行为识别已应用于智能视觉监控、人机交互、增强现实等领域,主要涉及底层特征的提取和表示以及模式的分类和识别两个阶段,它们构成了人体行为识别的重要步骤。现实条件中的部分遮挡、动态背景、移动摄像头、视角和光照变化等因素的影响,给人体行为识别带来了很大的挑战。从视频序列中提取出有效的运动特征是人体行为识别中重要的一环,能否提取出有效的运动特征,对行为识别的准确率和鲁棒性都有着直接的影响。文献[2]总结了常见的运动特征,将其大致分为4类:基于轮廓和形状的静态特征、基于光流或运动信息的动态特征、基于时空立方体的时空特征以及描述性特征,其中基于时空立方体的局部特征因其具有提取简单、抗干扰能力强、鲁棒性好等优点而日益成为新的研究热点。

受人类视觉系统稀疏性的启发,Olshausen和Field提出了稀疏表示模型,即信号可以由一个过完备的字典中的少部分原子线性组合而成,信号能量只集中于较少的原子,其中具有非零系数的原子揭示了信号的主要特征与内在结构。除了在图像恢复和压缩感知等领域取得了成功以外,最近的研究发现稀疏表示在模式分类上同样有着很好的效果,如人脸识别[5,11]、纹理识别[18-19]等。用于稀疏表示的字典通常需从样本中学习得到,目前的研究情况有:1)学习一个所有类别共享的通用字典,以及从系数中得到一个分类器来进行分类[5,17]。但是,通用的字典会丢失字典原子和类别标签之间的对应关系,因此不能通过每个类别的重构误差来进行分类。2)为每个类别的样本学习一个子字典,并保证各个类别的字典尽可能地独立,然后使用各个类别的子字典的重构误差作为鉴别信息来进行分类[3,19],但是并未使得稀疏编码系数具有鉴别性。文献[11]首次提出了Fisher鉴别字典学习方法,通过学习一个结构化的字典,并对编码系数施加Fisher鉴别准则,使得字典以及稀疏编码系数都具有鉴别性,同时通过实验验证了其在图像分类上的优越性能。文献[20]仅在编码系数上施加了Fisher鉴别准则,并用得到的具有鉴别性的稀疏编码系数作为特征来对人体行为进行识别,但是并未使字典具有鉴别性。

本文着重研究了基于Fisher鉴别字典学习方法的人体

到稿日期:2016-07-12 返修日期:2016-10-28 本文受江苏省科技支撑计划——社会发展项目:面向治安防控的监控视频目标检索关键技术研究(BE2014714),国家科技重大专项:无人装备智能控制支撑软件系统(2015ZX01041101)资助。

季 冲(1993-),男,硕士,主要研究方向为图像处理与模式识别,E-mail:jichong_1993@163.com;王 胜(1982-),男,博士,主要研究方向为模式识别,E-mail:wangsheng1910@163.com;陆建峰(1969-),男,教授,主要研究方向为图像处理与模式识别,E-mail:lujf@njust.edu.cn(通信作者)。

行为识别。首先在人体行为的视频序列中提取了 LMP(Local Motion Pattern)局部时空特征[3]，并用随机投影法[9]进行降维；然后对降维后的运动特征进行 Fisher 鉴别字典学习，该方法为每类样本学习一个子字典来得到一个结构化的字典，并在编码系数上施加了 Fisher 鉴别准则，使同一类样本的编码系数的离散度矩阵最小化而不同类的样本的编码系数的离散度矩阵最大化，从而同时增强字典和稀疏表示系数的鉴别能力；最后提出了一种新的分类方法，其同时利用重构误差和稀疏表示系数进行行为识别。实验结果验证了该方法在人体行为识别上的有效性和鲁棒性。

2　运动特征提取

本文采用文献[3]中提出的 LMP 运动特征描述子来对人体行为视频提取行为识别特征，该描述子针对基于一些二维的关键点检测器检测出的关键点(如 Harris 角点检测器等)的邻域提取时空立方体，捕捉它们在时间上的变化来作为运动特征。该描述子具有计算速度快、对时空尺度变化不敏感的特性。

给定一个 f 帧的视频序列 $V(x,y,t)$，将其分成 S 个时间片段，即 $V=[V_1,V_2,\cdots,V_S]$，即每个片段包含 $l=f/S$ 个连续的帧，l 表示时间上的分辨率，l 越小则时间分辨率越高。分辨率 l 不能太大也不能太小，要求能包含视频序列中的微小动作，但是不能包含大的改变。首先利用二维的关键点检测器对时间片段 $V_i(i=1,2,\cdots,S)$ 的第一帧检测出 ρ_i 个关键点；然后以每个关键点为中心从时间片段中提取一个大小为 $w\times w\times l$ 的时空立方体，w 为选取的邻域窗口的大小；接着根据该时空立方体，相对时间轴方向计算其二阶矩、三阶矩、四阶矩；最后将它们串联起来拉成一列向量，即得到 LMP 描述子。

图 1 给出了 LMP 描述子计算过程的一个示意图。

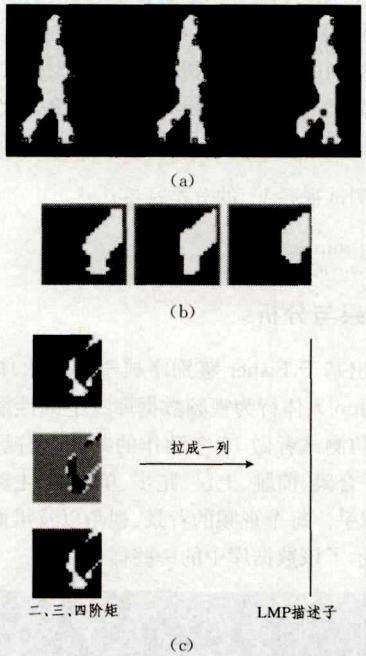


图 1　LMP 描述子计算过程示意图

图 1(a)展示了一个含有 3 个连续视频帧的时间片段，利用二维 Harris 关键点探测器对第一帧进行检测，并标记出关键点，在第二、三帧中对同样的位置进行了标记。图 1(b)展示了以某个关键点为中心对该长度为 3 帧的时间片段提取出的局部时空立方体。图 1(c)展示了对提取出的时空立方体沿时间方向计算其二、三、四阶矩，并将这 3 个矩的矩阵拉成一个列向量，即为 LMP 描述子。关于 LMP 描述子的更多细节参考文献[3]。

3　基于 Fisher 鉴别字典学习的人体行为识别

本节给出基于 Fisher 鉴别字典学习的人体行为识别的具体方法。

3.1　特征的维数约减

运动特征一般都具有高维度的特性，当选取邻域窗口的长度 $w=20$ 时，LMP 描述子的特征向量的维数为 1200×1，这样的高维度会影响计算速度以及实用性能，一般会采用标准的降维方法(如主成分分析(PCA)、线性鉴别分析(LDA)等)来去冗余，以获得高维数据在低维空间中的表示。本文采用了文献[9]中提出的随机投影法来对高维特征进行降维。理论研究表明，将数据投影到一个随机的低维子空间依然可以很好地保持向量之间的距离，且随机投影法十分简单、快速，这也是本文选择该降维方法的原因。

假定从一个视频序列中提取出了 n 个描述子，且每个描述子的维数为 p，用 $H\in\mathbb{R}^{p\times n}$ 来表示该描述子矩阵。随机投影法通过对描述子矩阵 H 左乘一个随机矩阵 $R\in\mathbb{R}^{m\times p}$ 来实现降维，将原始的 p 维描述子投影到一个 $m(m\leqslant p)$ 维子空间中，表达式如下：

$$L=RH \tag{1}$$

其中，随机矩阵 R 为零均值和单位方差的正态分布矩阵。

3.2　Fisher 鉴别字典学习方法

本节采用降维后的运动特征来学习一个结构化的字典，并基于 Fisher 鉴别准则得到具有鉴别性的稀疏表示系数。

用 $Y=[Y_1,Y_2,\cdots,Y_c]$ 来表示降维后的运动特征构成的训练样本集合，其中 Y_i 是第 i 类的样本。用 X 表示 Y 在 D 上的编码系数矩阵，一个字典学习问题可以表示为：

$$\min_{D,X}\{\|Y-DX\|_F^2+\lambda\|X\|_1\} \tag{2}$$

其中，$\|\cdot\|_F$ 代表矩阵的 Frobenius 范数，是对矩阵中所有元素的平方和的开根号；$\|\cdot\|_1$ 表示矩阵的 1 范数，是矩阵中所有列向量绝对值之和的最大值；λ 是一个可调标量参数。

为了提高字典学习的性能，本文采用基于 Fisher 鉴别的字典学习方法。不同于学习一个所有类别共有的字典，该方法是学习一个结构化的字典 $D=[D_1,D_2,\cdots,D_c]$，其中 D_i 是与第 i 类样本相对应的子字典，c 是类别的个数。类似 SRC 方法[5]，利用字典 D，可以用各个类的重构误差来进行分类。

把系数矩阵 X 写成 $X=[X_1,X_2,\cdots,X_c]$，其中 $X_i=\{x_j\}_{j=1}^{n_i}$ 表示第 i 类视频样本中提取的运动特征的集合 Y_i 在字典 D 上的编码表示系数，n_i 表示从第 i 类视频样本中提取出的特征描述子的个数。将 X_i 写成 $X_i=[X_i^1;X_i^2;\cdots;X_i^c]$，其中 X_i^j 表示 Y_i 在子字典 D_j 上的编码系数。为了保证字典 D 可以很好地重建样本集 Y，首先需要字典 D 能很好地表示 Y_i，因此有 $Y_i\approx DX_i=D_1X_i^1+\cdots+D_iX_i^i+\cdots+D_cX_i^c$。其次，由于 D_i 是与第 i 类对应的子字典，因此要求 D_i 能很好地表示 Y_i，而其他的 $D_j(j\neq i)$ 的表示能力略差。这意味着 X_i^i 应该具有一些显著系数，使得 $\|Y_i-D_iX_i^i\|_F^2$ 较小，而 X_i^j 的

系数应该接近于0，使得 $\|D_jX_i^j\|_F^2$ 较小。从而，可以定义如下准则来同时保证字典的表达能力和鉴别性：

$$R(Y_i,D,X_i)=\|Y_i-DX_i\|_F^2+\|Y_i-D_iX_i^i\|_F^2+\sum_{\substack{j=1\\j\neq i}}^{c}\|D_jX_i^j\|_F^2 \tag{3}$$

目前的字典学习方法很少考虑让稀疏编码系数具有鉴别性，通过引入 Fisher 鉴别准则可以使样本集 Y 在字典 D 上的编码系数 X 具有鉴别性。借鉴 Fisher 鉴别准则的思想，通过最小化 X 的类内离散度(记作 $S_W(X)$)并最大化 X 的类间离散度(记作 $S_B(X)$)来保证这一点。$S_W(X)$ 和 $S_B(X)$ 的定义如下：

$$\begin{aligned}S_W(X)&=\sum_{i=1}^{c}\sum_{k=1}^{n_i}(x_k^{(i)}-\mu_i)(x_k^{(i)}-\mu_i)^{\mathrm{T}}\\S_B(X)&=\sum_{i=1}^{c}n_i(\mu_i-\mu)(\mu_i-\mu)^{\mathrm{T}}\end{aligned} \tag{4}$$

其中，μ_i 和 μ 分别是第 i 类样本的编码系数矩阵 X_i 和总体编码系数矩阵 X 的中心，n_i 是 Y_i 中的样本个数。

直观上，可以最小化 $tr(S_W(X))-tr(S_B(X))$ 来作为优化目标，但这样是非凸和不稳定的。为了解决该问题，引入一个松弛项并定义如下准则来保证系数 X 的鉴别性：

$$F(X)=tr(S_W(X))-tr(S_B(X))+\eta\|X\|_F^2 \tag{5}$$

其中，η 是一个可调参数。

综上所述，Fisher 鉴别字典学习模型可以表示为：

$$J_{(D,X)}=\arg\min_{(D,X)}\{R(Y,D,X)+\lambda_1\|X\|_1+\lambda_2F(X)\} \tag{6}$$

其中，第一项 $R(Y,D,X)$ 用于保证字典 D 对样本集 Y 的表达能力和鉴别性，第二项 $\|X\|_1$ 是对系数的稀疏性约束，第三项 $F(X)$ 是对系数矩阵 X 施加的 Fisher 鉴别约束，λ_1 和 λ_2 是标量参数。

可以将 Fisher 鉴别字典学习模型的求解分解为两个子问题：固定字典 D 更新编码系数 X 和固定编码系数 X 更新字典 D。通过交替迭代求解来得到字典 D 和稀疏编码系数 X。

步骤 1　假设字典 D 是固定的，式(6)中的目标函数 $J_{(D,X)}$ 可简化为求解 $X=[X_1,X_2,\cdots,X_c]$ 的一个稀疏编码问题，这里逐个类别地求解 X_i。当求解 X_i 时，所有 $X_j(j\neq i)$ 是固定的。将式(6)中的目标函数进一步简化为：

$$J_{(X_i)}=\arg\min_{(X_i)}\{R(Y_i,D,X_i)+\lambda_1\|X_i\|_1+\lambda_2F(X_i)\} \tag{7}$$

且

$$F(X_i)=\|X_i-M_i\|_F^2-\sum_{k=1}^{c}\|M_k-M\|_F^2+\eta\|X_i\|_F^2 \tag{8}$$

其中，M_k 和 M 分别是第 k 类样本集和全体样本集的均值向量构成的矩阵(用 n_k 个均值向量 μ_k 和 μ 作为列向量)。可以用迭代投影法[21](IPM)求解式(7)。

步骤 2　假设编码系数 X 是固定的，当逐类地更新子字典 D_i 时，所有 $D_j(j\neq i)$ 是固定的。式(6)中的目标函数 $J_{(D,X)}$ 被简化为：

$$J_{(D_i)}=\arg\min_{(D_i)}\{\|Y-D_iX^i-\sum_{\substack{j=1\\j\neq i}}^{c}D_jX^j\|_F^2+\|Y_i-D_iX_i^i\|_F^2+\sum_{\substack{j=1\\j\neq i}}^{c}\|D_jX_i^j\|^2\} \tag{9}$$

式(9)是一个二次规划问题，其中 X^i 是 Y 在 D_i 上的编码系数。

3.3　分类器设计

当在学习得到字典 D 后测试样本即可在 D 上编码并进行分类。基于 Fisher 鉴别的字典学习算法，可以使重构误差和编码系数都具有鉴别性。自然地，可以同时利用它们来获得更准确的分类结果。

根据 3.2 节中求解出的系数矩阵 X(改写 $X=[X^1;X^2;\cdots;X^c]$，其中 X^i 表示与第 i 个类别的子字典相对应的稀疏表示系数矩阵)，计算与各个类别的子字典相对应的编码系数直方图，即

$$h_i=\frac{1}{n}\sum_{j=1}^{n}x_j^i,i=1,2,\cdots,c \tag{10}$$

h_i 即与第 i 个类别的子字典相对应的稀疏表示系数的中心，可以理解为原始信号转到稀疏域后的中心。其中，n 表示从训练集视频样本中提取的特征描述子的总个数，x_j^i 表示子矩阵 X^i 中的第 j 列向量。

给定一个测试视频 V_q，设提取出的运动特征经过维数约减后记为 $Y_q=\{f_j\}_{j=1}^{n_q}$(n_q 表示特征的个数)，记学到的字典 $D=[D_1,D_2,\cdots,D_c]$。因为训练集中各类的运动特征的数量相对较大，学到的子字典 D_i 可以很好地张成第 i 类的样本空间，所以可以直接在 D_i 上对 Y_q 进行编码来减小计算量以及其他子字典的干扰。求解式(11)得到 Y_q 在字典 D 上的稀疏表示系数矩阵，记为 $\hat{X}_q$：

$$\hat{X}_q=\arg\min_{X_q}\{\|Y_q-D_iX_q\|_F^2+\lambda\|X_q\|_1\} \tag{11}$$

其中，λ 是一个标量参数。以类似式(10)的方法计算 $\hat{X}_q$ 的系数直方图 h_q，最终将用来分类的度量定义为：

$$e_i=\|Y_q-D_i\hat{X}_q\|_F^2+\eta_1\|\hat{X}_q\|_1^2+\eta_2\|h_q-h_i\|_2^2 \tag{12}$$

其中，第一项为测试集在子字典 D_i 上的重构误差，第二项为对编码系数施加的稀疏性惩罚项，第三项为测试集在 D_i 上的编码系数矩阵的中心 h_q 与第 i 类子字典相应的编码系数的中心 h_i 之间的距离差；η_1 和 η_2 是可调的权重系数。最后，由式(13)得到测试视频 V_q 的分类标签 $\hat{i}_q$：

$$\hat{i}_q=\arg\min_{i\in[1,2,\cdots,c]}e_i \tag{13}$$

4　实验结果与分析

本节给出基于 Fisher 鉴别字典学习的人体行为识别方法在 Weizmann 人体行为视频数据库[8]上的性能表现。该数据库包含 9 位测试者做 10 个动作的共 90 个视频，行为类别包括弯腰、开合跳、前跳、上跳、跑步、单脚跳、走路、侧跑、单臂挥手、双臂挥手。每个视频的背景、视角以及摄像头都是固定的。图 2 展示了该数据库中的一些样例帧。



图 2　Weizmann 标准数据库中的一些示例帧

实验中的所有识别准确率都采用"留一法"方式得到，即

轮流选择其中一个人的行为视频序列作为测试样本，其余所有人的视频序列作为训练样本。在特征提取过程中，设 LMP 描述子的邻域窗口长度 w 为 20，对每个视频序列选取两个不同的时间分辨率，即时间片的个数 $S=[8,10]$。在字典学习过程中，设各个类别子字典原子个数为 120，分类时设权重参数 $\lambda=0.05$，$\eta_1=0.1$，$\eta_2=0.05$。图 3 给出了不同子字典原子下 Fisher 鉴别字典学习相比于传统 KSVD 方法的分类识别准确率，可以看到，相比于 KSVD 方法只能单纯使用重构误差进行分类，同时使用重构误差和编码系数的 Fisher 鉴别字典学习方法具有更高的行为识别性能。图 4 给出了识别结果的混淆矩阵，错分情况大多出现在跑步和走路、单手碰地面和单臂挥手这些比较近似的行为之间。表 1 列出了不同方法的行为识别准确率对比，可以看到本文方法基本上达到了目前的先进水平。

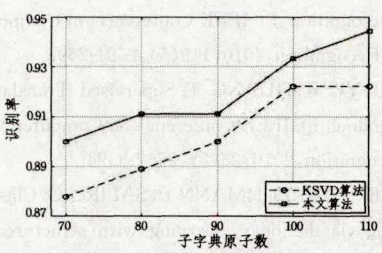


图 3　本文算法与 KSVD 算法的识别准确率比较

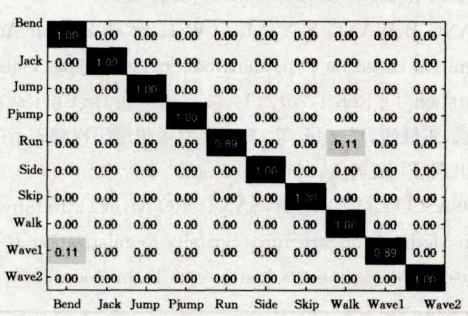


图 4　Weizmann 标准数据集上识别结果的混淆矩阵

表 1　Weizmann 数据集上不同方法的行为识别准确率对比

方法	准确率/%
Scovanner et al[12]	82.6
Klaser & Marszalek[13]	84.3
Niebles et al[16]	90.0
Zhang et al[14]	92.8
Seo et al[15]	97.5
Corelick et al[8]	97.8
本文方法	97.8

为了测试各种方法对视角变化、不规则行为方式的鲁棒性，Weizmann 鲁棒性数据库在 0°～81°之间的不同视角拍摄了 10 个测试视频，并在不同场景下拍摄了 10 组不同形态的行走视频。图 5 展示了不同形态下行走的一些样例帧，具体参见文献[8]。



图 5　Weizmann 鲁棒性数据库中的一些示例帧

本文将 Weizmann 标准数据库中的所有视频序列作为训练集，将鲁棒性视频序列作为测试样本，对 3 种分类方法进行了鲁棒性测试对照实验。测试结果如表 2 和表 3 所列。

表 2　Weizmann 鲁棒性数据集上 3 种方法对不同视角的行走动作的识别结果

不同视角	SRC	KSVD	本文方法
0°	Walk	Walk	Walk
9°	Walk	Walk	Walk
18°	Walk	Walk	Walk
27°	Walk	Walk	Walk
36°	Walk	Walk	Walk
45°	Walk	Walk	Walk
54°	Walk	Walk	Walk
63°	Jump	Run	Walk
72°	Jump	Jump	Walk
81°	Side	Walk	Jump

表 3　Weizmann 鲁棒性数据集上 3 种方法对不同形态的行走动作的识别结果

不同形态	SRC	KSVD	本文方法
Swinging a bag	Walk	Walk	Walk
Carry briefcase	Walk	Walk	Walk
Knees Up	Walk	Walk	Walk
Limping man	Walk	Walk	Walk
Sleep walking	Walk	Walk	Walk
Occluded Legs	Walk	Walk	Walk
Normal walk	Walk	Walk	Walk
Occluded by pole	Walk	Walk	Walk
Walking in a skirt	Walk	Walk	Walk
Walking with dog	Walk	Walk	Walk

从测试结果中可以发现，本文使用的行为识别方法对不同形态的变化不敏感，具有较高的鲁棒性。而在不同视角变换测试中可以看出，使用的分类方法能够容忍角度小于 72°的视角变化。综合来看，基于 Fisher 鉴别字典学习的人体行为识别方法对于视角变化、不同形态的变化等具有一定的鲁棒性。

结束语　人体行为识别是图像处理、计算机视觉、模式识别、机器学习、人工智能等多个学科的交叉研究课题，在视频监控、视频检索、人机交互、虚拟现实、医疗看护等领域具有重要的理论意义和很强的实用价值。本文根据人体行为识别的一般步骤，首先对人体行为视频序列提取了 LMP 局部时空特征并应用随机投影法进行降维，然后在对降维后的特征进行模式分类时引入了基于 Fisher 鉴别的字典学习方法，增强了字典和稀疏表示系数的鉴别能力，从而更有利于分类识别。实验结果验证了该分类方式在人体行为识别上的有效性和鲁棒性。

参 考 文 献

[1] POPPE R. A survey on vision-based human action recognition [J]. Image and Vision Computing, 2010, 28(6): 976-990.

[2] HU Q, QIN L, HUANG Q M. A Survey on Visual Human Action Recognition [J]. Chinese Journal of Computers, 2013, 36(12): 2512-2524. (in Chinese)
胡琼，秦磊，黄庆明. 基于视觉的人体行为识别综述[J]. 计算机学报，2013，36(12)：2512-2524.

[3] GUHA T, WARD R K. Learning sparse representations for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(8): 1576-1588.

[4] YUAN C, HU W, TIAN G, et al. Multi-task Sparse Learning with Beta Process Prior for Action Recognition[J]. IEEE Conference on Computer Vision & Pattern Recognition, 2013, 9(4): 423-429.

[5] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[6] AHARON M, ELAD M, BRUCKSTEIN A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.

[7] PATI Y C, REZAIIFAR R, KRISHNAPRASAD P S. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition[C]// Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers. 1993: 40-44.

[8] GORELICK L, BLANK M, SHECHTMAN E, et al. Actions as space-time shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2247-2253.

[9] BARANIUK R, WAKIN M. Random Projections of Smooth Manifolds[J]. Foundations of Computational Math, 2009, 9(1): 51-77.

[10] RUBINSTEIN R, ZIBULEVSKY M, ELAD M. Efficient Implementation of the K-SVD Algorithm Using Batch Orthogonal Matching Pursuit[R]. Haifa, Israel: Israel Institute of Technology, 2008.

[11] YANG M, ZHANG L, FENG X, et al. Fisher Discrimination Dictionary Learning for sparse representation[J]. International Conference on Computer Vision, 2011, 24(4): 543-550.

[12] SCOVANNER P, ALI S, SHAH M. A 3-dimensional sift descriptor and its application to action recognition[C]// Proceedings of the 15th international conference on Multimedia. ACM, 2007: 357-360.

[13] KLASER A, MARSZALEK M. A spatio-temporal descriptor based on 3D-gradients[C]// Proceedings of the British Machine Vision Conference. 2008: 995-1004.

[14] ZHANG Z, HU Y, CHAN S, et al. Motion Context: A New Representation for Human Action Recognition[C]// Proceedings of the 10th European Conference on Computer Vision. 2008: 817-829.

[15] SEO H J, MILANFAR P. Action recognition from one example [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 867-882.

[16] NIEBLES JC, WANG H, LI F F. Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.

[17] ZHANG Q, LI B. Discriminative K-SVD for dictionary learning in face recognition[J]. IEEE Conference on Computer Vision & Pattern Recognition, 2010, 119(5): 2691-2698.

[18] YANG J, YU K, HUANG T. Supervised Translation-Invariant Sparse coding[J]. IEEE Conference on Computer Vision & Pattern Recognition, 2010, 26(2): 3517-3524.

[19] RAMIREZ I, SPRECHMANN P, SAPIRO G. Classification and clustering via dictionary learning with structured incoherence and shared features[J]. IEEE Conference on Computer Vision& Pattern Recognition, 2010, 23(3): 3501-3508.

[20] WANG B, WANG Y Y, XIAO W H, et al. Human Action Recognition Based on Discriminative Sparse Coding Video Representation[J]. Robot, 2012, 34(6): 745-750. (in Chinese)
王斌, 王媛媛, 肖文华, 等. 基于鉴别稀疏编码视频表示的人体动作识别[J]. 机器人, 2012, 34(6): 745-750.

[21] ROSASCO L, VERRI A, SANTORO M, et al. Iterative Projection Methods for Structured Sparsity Regularization[R]. Massachusetts: Massachusetts Institute of Technology, 2009.

(上接第 256 页)

[9] BALTAG A, MOSS L, SOLECKI S. The logic of public announcements, common knowledge, and private suspicions[C]// Proceedings of the Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98). 1998.

[10] VAN BENTHEM J, VAN EIJCK J, KOOI B. Logics of communication and change[J]. Information and Computation, 2006, 204(11): 1620-1662.

[11] LÖWE B, PACUIT E, WITZEL A. DEL Planning and Some Tractable Cases[C]// LORI. 2011: 179-192.

[12] BOLANDER T, ANDERSEN M B. Epistemic planning for single-and multi-agent systems[J]. Journal of Applied Non-Classical Logics, 2011, 21: 9-34.

[13] VAN EIJCK J. DEMO-A Demo of Epistemic Modelling[M]// Prceedings of the 7th Augustus de Morgan Workshop. Amsterdam. 2007: 305-363.

[14] YU Q, WEN X M, LIU Y M. Multi-agent epistemic explanatory diagnosis via reasoning about actions [C]// Proceedings of the Twenty-Third International Joint Conference on Articial Intelligence (IJCAI-13). 2013

[15] AUCHER G, BOLANDER T. Undecidability in Epistemic Planning [C]// Proceedings of the Twenty-Third International Joint Conference on Articial Intelligence (IJCAI-13). 2013

[16] BOLANDER T, JENSEN M H, SCHWARZENTRUBER F. Complexity Results in Epistemic Planning [C]// Proceedings of the 24rd International Joint Conference on Artificial Intelligence(IJCAI-15). Buenos Aires, Argentina, 2015: 3-9.

[17] MUISE C J, BELLE V, FELLI P. et al. Planning Over Multi-Agent Epistemic States: A Classical Planning Approach [C]// Proceedings of the Twenty-Ninth Conference on Artificial Intelligence(AAAI-15). Austin, Texas, USA, 2015: 3327-3334.

[18] YANG R, WAN H, FANG L, et al. A Complete Epistemic Planner without Epistemic Closed World Assumption [C]// Proceedings of the 24rd International Joint Conference on Artificial Intelligence(IJCAI-15). Buenos Aires, Argentina, 2015: 3-9.