



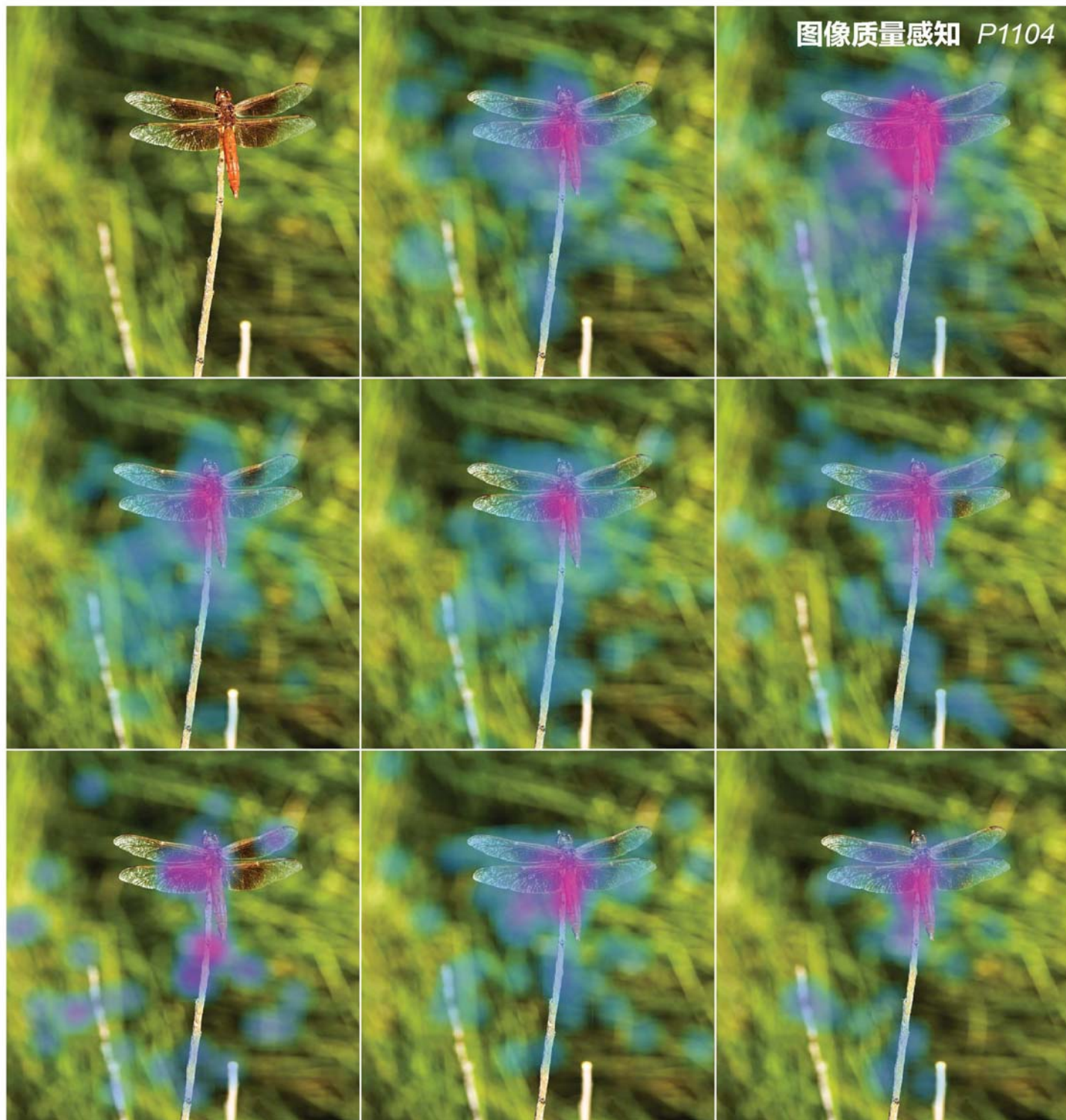
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

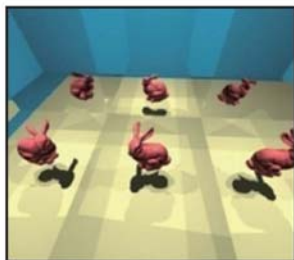
2014
07
VOL.19

ISSN1006-8961
CN11-3758/TB



多特征融合的图像自动变形
(第1012页)

多通道Haar-like特征多示例学习目标跟踪(第1038页)

快速虚顶点光子映射的焦散
绘制(第1062页)**综述**

打印机光谱特性化方法研究综述

万晓霞, 刘强 0985

图像处理和编码

图像场方向导数的局部区域重建

叶学义, 齐珍珍, 何志伟, 李汪兵 0998

Lytro相机的光场图像校正与重对焦方法

周文晖, 林丽莉 1006

多特征融合的图像自动变形

肖红光, 肖树根, 刘京, 李淳芃, 陈立福 1012

结合频谱扩散的指纹图像滤波增强

卞维新, 封举富, 罗永龙, 徐德琴 1021

图像分析和识别

中心投影变换的抗噪性分析

黄永东, 李三三, 杨建伟 1031

多通道Haar-like特征多示例学习目标跟踪

宁纪锋, 赵耀博, 石武祯 1038

自主移动机器人自适应室外道路检测

杜明芳, 王军政, 李静, 崔广涛, 方建军, 曹海青 1046

分层树结构字典编码的行为识别

周同驰, 程旭, 吴镇扬 1054

计算机图形学

快速虚顶点光子映射的焦散绘制

李瑞瑞, 秦开怀 1062

双重视图下的3维网格模型可见水印

陆尘, 朱长青, 王玉海 1068

虚拟现实与增强现实

个性化编辑的轻量化3维树木模型构建

董天阳, 纪磊, 刘思远, 范菁, 熊丽荣 1074

遥感图像处理

带距离徙动校正的压缩感知PFA成像方法

段化军, 朱岱寅 1085

结合边缘和区域的活动轮廓模型SAR图像目标轮廓提取

王沛, 周鑫, 彭荣鲲, 符鹏 1095

第2届全国图像图形联合学术会议专栏

视线跟踪技术的图像质量主观数据库构建及其可视化

张昀, 牟轩沁 1104

丰富图像标签的正则化非负矩阵分解方法

夏召强, 冯晓毅, 彭进业 1112

CONTENTS

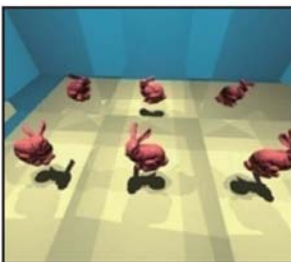
JOURNAL OF IMAGE AND GRAPHICS



Multi-feature fusion-based image morphing(P1012)



Multiple instance learning based object tracking with multi-channel Haar-like feature(P1038)



Fast caustics rendering by photon beam tracing based on virtual vertex transformations(P1062)

Review

- Review of spectral printer characterization
Wan Xiaoxia, Liu Qiang 0985

Image Processing and Coding

- Reconstructing image local regions based on directional derivative of a field
Ye Xueyi, Qi Zhenzhen, He Zhiwei, Li Wangbing 0998
- Light field image rectification and refocus method for Lytro camera
Zhou Wenhui, Lin Lili 1006
- Multi-feature fusion-based image morphing
Xiao Hongguang, Xiao Shugen, Liu Jing, Li Chunpeng, Chen Lifu 1012
- The fingerprint enhancement using filtering combined with spectrum diffusion
Bian Weixin, Feng Jufu, Luo Yonglong, Xu Deqin 1021

Image Analysis and Recognition

- Robustness to noise of central projection transformation
Huang Yongdong, Li Sansan, Yang Jianwei 1031
- Multiple instance learning based object tracking with multi-channel Haar-like feature
Ning Jifeng, Zhao Yaobo, Shi Wuzhen 1038
- Adaptive outdoor road detection for autonomous mobile robot
Du Mingfang, Wang Junzheng, Li Jing, Cui Guangtao, Fang Jianjun, Cao Haiqing 1046
- Action recognition using hierarchically tree-structured dictionary encoding
Zhou Tongchi, Cheng Xu, Wu Zhenyang 1054

Computer Graphics

- Fast caustics rendering by photon beam tracing based on virtual vertex transformations
Li Ruirui, Qin Kaihuai 1062
- Visible watermarking for three-dimensional mesh model in two views
Lu Chen, Zhu Changqing, Wang Yuhai 1068

Virtual Reality and Augmented Reality

- Construction of lightweight three-dimensional tree models supporting personalization editing
Dong Tianyang, Ji Lei, Liu Siyuan, Fan Jing, Xiong Lirong 1074

Remote Sensing Image Processing

- PFA imaging with range migration correction by compressive sensing
Duan Huajun, Zhu Daiyin 1085
- Active contour model based on edge and region attributes for target contour extraction in SAR image
Wang Pei, Zhou Xin, Peng Rongkun, Fu Peng 1095

Column of UCIG'2013

- Eye tracking based image quality assessment database and visualization method
Zhang Yun, Mou Xuanqin 1104
- Regularized non-negative matrix factorization method for image tag enrichment
Xia Zhaoqiang, Feng Xiaoyi, Peng Jinye 1112

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2014)07-1054-08

论文引用格式: Zhou T C, Cheng X, Wu Z Y. Action recognition using hierarchically tree-structured dictionary encoding[J]. Journal of Image and Graphics, 2014, 19(7): 1054-1061. [周同驰, 程旭, 吴镇扬. 分层树结构字典编码的行为识别[J]. 中国图象图形学报, 2014, 19(7): 1054-1061.]
[DOI:10.11834/jig.20140709]

分层树结构字典编码的行为识别

周同驰, 程旭, 吴镇扬

东南大学信息科学与工程学院, 南京 210096

摘要: **目的** 基于学习字典的稀疏编码能够自适应地表示信号。然而,传统学习字典的原子之间缺少关联,信号的相似性在编码后缺失。考虑到结构化稀疏表示的鲁棒性和判别性能力,结构化字典的构建成为一个重要的任务。**方法** 依据标准的凸优化字典学习算法,引入数据点编码路径的约束(由上层原子激活的索引规划下层的索引),构思了一种树结构字典学习框架。**结果** 实验结果表明,局部描述符的稀疏表示具有较好的鲁棒性和判别性,同时在 KTH 数据库上人体行为识别实验与其他类似文献方法相比获得了较高的识别精度,其中,时空梯度方向直方图(HOG3D)的编码识别结果达到 97.99%。**结论** 通过实验结果,观察到采用本文构建的字典编码信号具有较好的鲁棒性和判别性,更好的适合分类任务。

关键词: 行为识别;稀疏编码;局部描述符;树结构字典

Action recognition using hierarchically tree-structured dictionary encoding

Zhou Tongchi, Cheng Xu, Wu Zhenyang

School of Information Science and Engineering, Southeast University, Nanjing 210096, China

Abstract: **Objective** Sparse coding using learned dictionaries can adaptively represent signals. However, the similarity among the signals that are encoded in traditional dictionary are lost due to a lack of correlations between atoms. Considering the robustness and discriminative power of structured sparse representation, the building of the structured dictionary becomes an important task. **Method** We conceive a framework of tree-structured dictionary by introducing a constraint for the data point code path (programming the index from the upper layer to the next layer) according to the standard convex optimization dictionary-learning algorithm. **Result** Experimental results on the KTH human action database show that local descriptor codes with learned tree-structured dictionary have good robustness and discriminative and demonstrate that our algorithm generally obtains higher recognition accuracy than other similar methods. We achieve an accuracy rate of 97.99% using histograms of oriented 3D spatial-temporal gradients (HOG3D). **Conclusion** From our experiments, we observe that the encoding of signals using the constructed dictionary has good robustness and discrimination, and is preferable for the task of classification.

Key words: action recognition; sparse coding; local descriptor; tree-structured dictionary

收稿日期:2013-11-13;修回日期:2014-01-21

基金项目:国家自然科学基金项目(60971098,60672094)

第一作者简介:周同驰(1979—),男,现为东南大学信息与通信工程专业博士研究生,主要研究方向为视频人体行为识别。

E-mail: zhoutongchi@sina.com

0 引言

采用传统的方法对海量的图像或视频数据处理、分析、理解、存储和传输已不能满足当今需求,如何将这庞大的数据实现简洁、自适应的表达,同时消除人类理解与底层数据之间的鸿沟成为人们日益关注的问题之一。近年来,海量数据的稀疏表示已有较好的发展,表现出了重要的研究意义,信号的稀疏表示在图像去噪、图像或视频的检索、视频行为识别等实际应用领域越来越受到人们的重视,其中基于结构化的稀疏表示逐渐成为人们研究的热点之一。对于传统无结构的稀疏编码^[1]缺少鲁棒性,即相似的特征在编码后缺失相似性。一些研究者提出群(组)结构^[2-3]、图或超图嵌入模型^[4]、树状结构^[5-6]等稀疏编码模型。这些方法将信号编码为具有鲁棒性和判别性的稀疏表示,在图像去噪、图像增强、分割、行为识别等应用方面已经表现出较好的性能。

2006年Lee等人^[1]提出了松弛 L_0 范数的凸优化稀疏编码和字典学习算法,迭代解决两个凸优化问题(L_1 规则和 L_2 约束最小二乘问题),明显地加快了编码和字典更新过程。文献[2]提出了群(组)稀疏的结构字典学习模型,结合凸分析和单调算子理论优化学习结构字典,算法具有收敛速度快,训练的字典能够更好地适应数据,大大提高了图像增强的效果。2013年Gao等人^[4]提出了图或超图嵌入规则化的稀疏编码,组织字典原子之间的关系,编码系数具有很好鲁棒性,该方法在图像分类、分割、图像标注等方面表现出较好的性能。2013年Wang等人^[6]采用复合惩罚规则和K奇异值分解(K-SVD)序贯的更新字典的方法交替优化学习组或树状结构字典,利用结构字典获得了较好的图像降噪效果。2010年Zhu等人^[7]提出了迁移学习构造字典方法,3D时空特征描述符的稀疏编码通过最大池运算表示行为,在KTH(瑞典皇家工学院)等人体行为视频数据库上获得了较高的识别精度。2011年Le等人^[8]提出了分层学习字典的方法,先无监督学习小尺度视频块的时空特征,之后利用学习的滤波器卷积学习大尺度的时空特征,同样在行为识别领域获得了较高的识别精度。

依据前研究者的理论基础,分析了 L_0 范数、 L_1 范数编码和字典更新过程,并研究了当前构建结构

字典的方法,提出了一种分层树结构字典学习框架。在该框架下,序贯分层学习多个子字典,通过上层选择字典原子的索引,规划数据点在下层选择的路径和原子索引,构建层与层之间的字典原子存在关联,层内原子分组的分层树结构字典。通过实验,信号的稀疏表示具有较好的鲁棒性和判别性,与同类文献方法相比,本文基于局部特征描述符的稀疏表示和词汇包(bag-of-words)模型表示视频的人体行为识别实验获得了较高的识别精度。

1 相关文献与存在的问题

在信号的稀疏表示中,字典的学习扮演着重要的角色,字典的更新和信号的稀疏编码是两个交替优化的过程。信号的近似逼近主要有贪婪追踪、凸松弛、贝叶斯框架等。目前,稀疏编码主要以前两者为主,基于匹配追踪(MP)和正交匹配追踪(OMP)等稀疏编码的非凸优化贪婪算法是一种逐步选择数据点与一个原子最佳匹配的过程,数据是基于匹配原子计算后的信号残差,通常采用K-SVD序贯地更新字典原子,而 L_1 范数的稀疏编码过程与此有较大的差别。传统的 L_1 范数规则化的编码过程是一个求解凸优化的问题,将信号的稀疏表示问题定义为一类有约束的极值问题,进而转化为线性规划问题来求解,编码过程协作考虑各原子对数据点的贡献,每次匹配一个原子后,更新编码系数,字典的更新通常采用拉格朗日对偶算法或采用在线更新策略。这些字典学习方法是假定数据之间相互独立的,仅考虑编码的稀疏度和残差的剩余,缺少码系数的先验信息或原子之间的关系。当信号微小变化时,致使编码系数有较大的差异性,相似的特征失去了相似的编码性质^[4,5,9]。标准的编码方法已不能满足高要求的模式分类、图像去噪、分割等领域的需求,而基于结构字典学习的编码算法能够获取鲁棒性的稀疏码,成为当前的研究热点,尤其是组结构、图嵌入结构、树状结构等的字典学习。

2012年,Zelnik-Manor等人^[10]在K-SVD算法^[11]基础上研究了块结构字典,块作为字典原子的约束。块结构字典对信号的稀疏表示具有鲁棒性和判别性,不足之处,块数目及每块的大小为固定值,块与块之间、块内原子缺少关联。2012年,Li等人^[3]在2006年Aharon等人^[11]提出的OMP编码和K-SVD字典更新算法的基础上,构建了原子的图模

型,学习图嵌入模型的组结构字典,有效地保留了图像块的几何结构^[4]信息。标准的 MP 或 OMP 算法是一个非凸优化和 NP (non-deterministic polynomial) 难问题,2006 年, Lee 等人^[1] 提出松弛 L_0 范数的方法来优化学习字典,其字典学习模型为

$$\begin{aligned} D = \arg \min_{D, \Phi} \|X - D\Phi\|_F^2 + \lambda \|\Phi\|_1 \\ \text{s. t. } \|\mathbf{D}_j\|_2^2 \leq 1, j = 1, 2, 3, \dots, n \end{aligned} \quad (1)$$

式中, X 为数据矩阵, 每个向量代表一个数据点, $\mathbf{D}_j$ 为字典 D 索引号为 j 的一个原子, Φ 为 X 的编码系数矩阵, F 表示 Frobenius 范数, λ 为平衡参数, 通常为经验值, 用来平衡稀疏表示的稀疏度与数据点的误差。2010 年 Jenatton 等人^[5] 在研究标准 L_1 规则稀疏编码基础上, 定义了树状结构字典, 将 L_1 范数嵌入到树结构分层范数, 采用混合范数作为规则项, 通过近邻梯度法优化学习结构化稀疏编码和坐标下降算法在线更新学习树状结构字典, 稀疏编码模型为

$$\begin{aligned} \min_{D \in \mathcal{D}^1, \Phi \in \mathbb{R}^{num \times M}} \frac{1}{M} \sum_{i=1}^M \left[\frac{1}{2} \|X_i - D\Phi_i\|_2^2 + \lambda \Omega(\Phi_i) \right] \\ \Omega(\Phi_i) = \sum_{g \in \zeta} w_g \|\Phi_{i, |g}\|_p \end{aligned} \quad (2)$$

式中, Ω 是规则项, 通常为混合范数, w_g 为正权重, p 通常采用 L_2 或 L_∞ 范数, D 为字典, $\mathcal{D}^1$ 表示凸集合, Φ_i 为第 i 个数据点 X_i 的稀疏编码, num 表示编码向量长度, M 为样本个数, ζ 表示组的集合。当 Ω 为 L_1 范数函数时, 式(2)与 Lee 等人^[1] 提出的 L_1 范数规则化编码等价。该字典学习是求解一个大的优化过程, 在线更新字典原子的同时需要依据定义树结构分配原子归属哪个结点, 计算的复杂度较高。2012 年 Guo 等人^[2] 提出了群稀疏的结构字典学习数学模型, 引入定位因子结合凸分析和单调算子理论提出了一种有效的结构化字典学习算法, 但是群的大小及字典原子的更新时效性仍然是需要解决的问题。组结构^[2-3]是从平均最大化相关性的角度编码数据点, 并非是一种很好的先验信息, 在学习组结构稀疏码或字典的同时, 组内非最佳原子抑制了数据点的显著信息。2012 年 Wei 等人^[9] 在标准的 L_1 范数规则化稀疏码的基础上, 采用了权重稀疏码规则化, 即引入了位置规则项, 计算数据点与字典原子之间的相似度作为编码系数的权重, 此种方法与图嵌入稀疏编码^[4]类似, 但又有所不同, 图嵌入模型并未考虑字典原子与数据点之间的局部位置, 数据点的稀疏编码具有鲁棒性, 但图嵌入方法在分类工

作中并不是最优越的编码方法^[9]。

2 结构字典学习及局部特征描述

L_0 范数的求解速度较快, 但缺乏稳定性, 而基于 L_1 范数的稀疏编码对不同特征具有较好的区分性, 对于相似的特征的编码缺少鲁棒性^[1,12]。引入组、块、树结构等先验信息的稀疏编码采用较少的原子逼近原始信号, 具有更好的鲁棒性和判别性, 但采用凸优化学习一个大的字典带来较大的计算复杂性。将一个大的字典学习优化过程分解为由粗到细关联^[13]的多个子字典序贯的学习过程, 通过平衡参数等条件来约束上层学习低频特征、下层学习高频特征的字字典, 层与层之间的原子存在关联, 层内原子分组。考虑到字典的原子具有复共用现象和相似信号选择原子的差异性, 定义一种分层框架学习字典的数学模型为

$$\begin{aligned} D_l = \arg \min_{D_l, \Phi_l} \|X_l - D_l\Phi_l\|_F^2 + \lambda_l \Omega(\Phi_l) \\ \text{s. t. } \|\mathbf{D}_{l,j}\|_2^2 \leq 1, j = 1, 2, 3, \dots, n \\ X_l = X_{l-1} - D_{l-1}\Phi_{l-1}, l = 1, 2, 3, \dots \end{aligned} \quad (3)$$

式中, D_l 表示第 l 层的子字典, Φ_l 表示数据 X_l 的编码系数矩阵, 每个列向量表示一个数据点的稀疏表示, λ_l 表示 l 层的平衡参数, $\mathbf{D}_{l,j}$ 为字典 D_l 索引号为 j 的一个原子。对于某一层, 采用 Lee 等人^[1] 提出的松弛 L_0 范数稀疏编码和拉格朗日对偶更新字典, 依据编码系数更新激活矩阵(数据点与原子的关联矩阵, 可由编码系数决定)。对于任一数据点, 通过数据点与上层的字典原子的关系(激活矩阵)预先规划下层可预选编码路径及路径内预选激活原子, 由上层激活原子的索引推导下层预选激活原子的索引, 其推导过程为

$$\begin{aligned} d_{l,j} \downarrow \\ d_l \in \{d_{l,i} \mid i = (S-1) \times n + j, S = 1, 2, 3, \dots, N\} \\ \text{s. t. } i \leq N \times n, N = 2, 3, \dots \end{aligned} \quad (4)$$

式中, n 为 l_1 层字典原子的数目, N 为 1 层与 l_1 层之间的字典大小关系, $d_{l,j}$ 表示 l_1 层索引号为 j 的原子, d_l 是由 $d_{l,i}$ 构成的一个子集, 索引号为 i 的原子是由式(4)推导而来(向下箭头表示推导关系), i 为一变量, S 为从 1 到 N 逐步取整数值, 若某个信号激活了上层第 j 个原子, 则下层激活的原子由 d_l 子集构成。随着分层学习字典的层数增加, 考虑到相似

的数据点选择的原子将呈现多样性,定义下层字典原子的数目是上一层的整数倍。

2.1 结构字典学习算法

在参考其他树模型定义的规则基础上,本文的分层树状字典学习模型规则定义如下:

- 1) 高层节点作为下一层的父节点,则其有1个或多个原子组成一个节点;
- 2) 在每层内,每个数据点可以有1个或多个父节点;
- 3) 父节点非空时,有可能产生子节点,组内可以产生多个节点;
- 4) 在学习过程中,字典的学习是自上向下分层学习,数据信息自上向下传递。

每层字典是欠完备的,由于 L_1 范数的凸优化的性质,子字典是一个判别性的子字典。定义的分层树状、分组字典学习过程的模型图如图1所示。

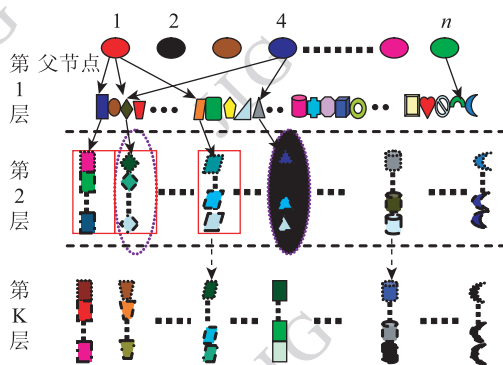


图1 分层结构字典学习模型

Fig. 1 Model of hierarchically structured dictionary

图1中,椭圆形的标示代表数据点,数据点1在第1层由3个原子(不同形状及亮度的图示代表一个原子)组成一个父节点,数据点4在第1层有2个原子,其中一个原子与数据点1共享,第 n 个数据点由1个原子构成父节点。从节点内的原子共享可以看出,数据点1与4有某些相似性。根据上层选择原子的索引,规划数据点在下层选择原子的索引和编码路径,数据点1在第1层由3个原子构建的一个父节点,预定义在第2层产生3个父节点(相对上层为子节点,相对于下层为父节点),数据点4与数据点1在第1层共享1个原子,由于上层共享的原子,数据点1、4在第2层的共享组内优化学习各自的父节点或共同父节点。通过以上分析,总结本文的训练字典算法如下:

- 1) 根据构思的字典学习框架模型,首先初始化

参数:

第1层:规则化参数 λ_1 ,随机初始化 num_1 个原子的子字典 D_1 (为加速收敛,通常随机选择部分训练样本作为字典原子),数据点激活索引矩阵 Act_1 (数据点与原子之间关联的共生矩阵即稀疏模式,首层初始化为0矩阵,大小为 $num_1 \times n$, n 为数据点个数);

第2层:规则化参数 λ_2 ,层之间倍数关系 N ,子字典原子个数为 $num_2 = N \times num_1$, $N \geq 2, 3, 4, \dots$,在这里选择 $N=2$,即当前层原子数是上一层的2倍,表示一个原子的父节点将会产生2个单原子的子节点。初始化字典 D_2 ,通常情况下第2层子字典也是随机初始化,为加快训练,采用复制第1层训练的子字典原子。第3层...

2) 以Lee等人^[1]算法为基础,学习第1层字典,满足条件时停止迭代优化,学习字典 D_1 作为第1层训练字典,记录每个数据点激活的原子索引,构建数据点与原子之间的激活索引矩阵,更新 Act_1 (原子激活,对应的激活矩阵元素位置置1,未激活置0)。

3) 第1层重构后的信号残差作为第2层学习字典的输入信号。由激活矩阵 Act_1 根据式(4)预先规划第2层激活矩阵 Act_2 作为初始激活矩阵。若第1层激活的索引号为 j ,则第2层预定义的激活索引号为 $(S-1) \times num_1 + j$, $S=1, 2, 3, 4, \dots, N$, N 为层与层的倍数关系, S 为从1到 N 逐步取整数值,当 $N=2$ 时, S 分别取1和2。

4) 在预激活矩阵约束编码路径的条件下,根据Lee等人^[1]算法,用初始化的子字典 D_2 对信号编码,达到条件时,停止迭代交替优化, D_2 作为第2层训练子字典,依据编码系数更新激活矩阵 Act_2 。

- 5) 若训练的层数未达到,重复步骤3)4)。

该框架训练的字典具有分层树状组结构,层与层之间的流程类似于 L_0 范数优化编码过程,层内考虑到信号的显著几何结构的多样性,采用 L_1 规则凸优化的编码算法。该算法考虑了层与层之间原子的父子节点关系,在层内采用 L_1 规则稀疏编码改善了文献[2-3]组稀疏编码不足。采用图像块信号来学习一个3层字典,学习的字典如图2所示,每层为一个子字典,每个子字典的列为组,下层原子的索引号与上层对应关联,组内原子存在共用现象。随着层数的增加,字典原子的特征由低频(模糊特征)到高频(细节特征)变化。

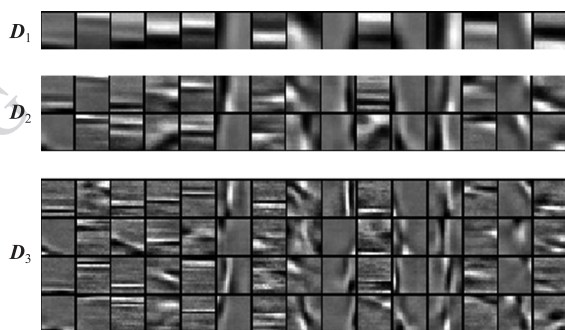


图2 分层字典

Fig. 2 Hierarchical dictionary

2.2 局部描述符的结构稀疏编码

视频或图像序列时空特征描述符主要有3DSIFT、HOG3D等,这些由2D拓展到3D的局部时空特征描述符在人体行为识别中被广泛的应用。HOG3D、云兴趣点^[14]等描述符融入了空间、时间的局部运动信息,具有较好的判别性能。基于稀疏编

码的行为识别主要从编码鲁棒性的描述符和编码时空上下文关联的局部特征(如融入几何结构的对特征等)等方面考虑判别性的表示行为,Zhu等人^[7]在迁移学习的指导下学习字典,以3D局部特征的鲁棒性稀疏编码作为行为单元;Zhang等人^[15]在密集光流域计算位移信息,采用稀疏码描述上下文内容;Zhang等人^[16]考虑了时空兴趣点的上下文关系,采用上下文约束的稀疏编码方法表示行为。为验证分层树状结构字典的有效性,将训练字典框架应用到行为识别领域,分别采用HOG3D^[7]、三平面正交局部二值模式(LBP-TOP)^[17]及随机采样视频块^[8]作为信号,编码局部特征描述符。首先,验证在该框架下信号的稀疏编码具有较好的鲁棒性和判别性,学习3层字典,如图2所示。对3个图像块信号编码如图3(a)~(c)所示,3个编码系数图是其右下角图像块的编码表示,图像块与图3(d)左边一列对应。

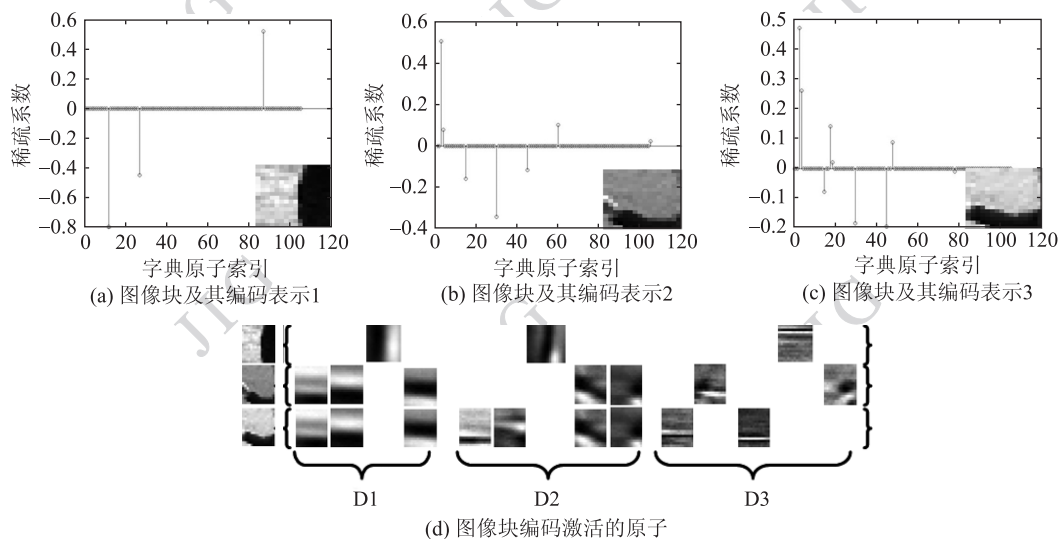


图3 图像块的编码表示和原子的选择

Fig. 3 Image patches coding representation and choice of atoms

图3(a)是从Patch集中选择差异较大的一个Patch(该图右下角的图像块)的编码系数,图3(b)(c)为3D-Patch中相邻图像块的编码系数,对应的原始信号为图3(d)左列下两行的图像块(与稀疏系数图示右下角的图像块相同),图3(d)括号内为原始信号选择的原子及其所在的子字典D1、D2、D3。从图3(b)(c)和图3(d)后两行可以看出,相邻图像块之间有较强的相似性,信号的稀疏表示也具有很强的相似性,尤其是在第1层,字典复共用的

原子完全相同,这就说明它们有共同的父节点,第2层存在部分复共用现象,随着层数增加,复共用现象逐渐变弱,选择的原子具有多样性,例如第3层,不存在共用的原子。图3(a)与图3(b)(c)的编码系数具有较大的差异性(判别性),稀疏非0系数选择的原子差异性较大,从图3(d)第1列中可以看出第1个图像块与后两个具有较大的差异性(几何结构差异大),稀疏表示时选择的原子完全不同。训练的分层字典对每个信号可以由较少的原子组合逼

近,分层结构的稀疏码表示具有较好的鲁棒性和判别性。

对图像块稀疏编码,如何用图像块的稀疏表示来描述视频块的3D特征呢? Thornton等人^[18]采用了时间池对图像序列的滤波响应堆栈处理来描述视频块特征。基于稀疏表示的图像分类、人体行为识别^[7]等应用领域,空域最大池的操作能很好地产生图像块或视频块的描述符,它具有较好判别性。本文将空间最大池的操作应用到图像块序列的稀疏表示上,时间尺度上形成的特征矩阵采用最大池操作产生视频块的稀疏表示,对某个原子采用最强的响应表示视频块的显著或局部微小变化特征,时间池的操作能够捕获局部微小运动信息和视频块的显著特征,即

$$V_j = \max \{ |\Phi_{j,1}|, |\Phi_{j,2}|, \dots, |\Phi_{j,N_{temp}}| \} \quad (5)$$

式中, V_j 为第 j 个剪辑视频块以行向量形式的稀疏表示, $\Phi_{j,i}$ 代表第 j 个视频块中第 i 个图像块的稀疏编码, $\Phi_{j,i}$ 为一行向量, N_{temp} 为视频块的时域长度,在这里,由于字典学习原子较少,不易在时域采用较大的尺度。

3 仿真实验及结果分析

采用KTH视频数据库验证上述基于结构字典的稀疏表示在行为识别应用领域的有效性。实验平台为i5 760台式机,四核处理器,主频为2.8 G,内存4 G,64位操作系统。采用最常用、最有效的行为表示及识别模型,即BoWs统计模型表示行为。最近邻域的NNC($k=3$)分类器,留一法评价分类识别效果,行为表示及识别通道类似Bregonzio等人^[14]的方法。由于随机采样视频块的数据量大,对所有局部时空特征的稀疏表示进行降维、量化,之后采用特征选择的方法选择判别性的描述符串联表示视频行为。在实验中,参考时空兴趣点提取空时体的尺度,不同的3D特征描述符采用不同的时间尺度,HOG3D选择大的时间尺度将引入更多的时空信息。为了避免学习字典的过匹配问题,逐步增加层数及初始第1层字典原子数,经过多次实验确定字典的层数及第1层原子的个数,训练字典所需数据的参数设置为:从599个视频切片中随机选择 10^2 个视频序列作为训练数据,共 10^5 个数据点。

1) 3D-Patch 时空尺度为 $16 \times 16 \times 7$,定义训练

字典参数:4层字典,每层原子数为 $num_1 = 10$, $num_2 = 20$, $num_3 = 40$, $num_4 = 80$,共150个原子的字典,经验平衡参数 λ 分别为:0.2,0.15,0.1,0.08;

2) HOG3D 空时尺度为 $15 \times 15 \times 10$ (支撑区域),分割积分视频块为 $3 \times 3 \times 2$ 细胞单元,定义训练字典参数:3层字典,每层原子数为 $num_1 = 15$, $num_2 = 30$, $num_3 = 60$,共105个原子的字典,经验平衡参数 λ 分别为0.15,0.1,0.08;

3) LBP-TOP 空时尺度为 $16 \times 16 \times 7$,LBP-TOP描述符是采用统一模式的8邻域计算直方图的方法,定义训练字典参数:3层字典,每层原子数为 $num_1 = 15$, $num_2 = 30$, $num_3 = 60$,共105个原子的字典,经验平衡参数 λ 分别为0.15,0.1,0.08。

对3D局部特征的3种描述符的稀疏表示采用主成分分析(PCA)降到稀疏表示信号长度的2/5,线性10个量化区间量化码系数,采用统计直方图作为视频块的描述符^[14-15],根据本文数据块的采样密集度,采用特征选择方法从统计直方图中选择部分特征描述符,串联作为视频切片的表示。

3.1 实验数据

KTH数据库是瑞典皇家工业大学CVAP实验室采集整理而成的视频数据,共599个视频切片,在4种条件下,采集25个人的6种行为,图4给出了KTH数据库内6类行为分别为box、clap、wave、jog、run、walk示例图像,每一行图像样本对应4种不同条件下的一种行为。



图4 KTH数据库部分示例图像

Fig. 4 Sample frames from KTH

3.2 实验结果及分析

分别计算HOG3D、LBP-TOP描述符和基于灰度

值像素的视频块的稀疏表示。对于 3D 图像块稀疏表示,首先对 3D-Patch 的每个图像块稀疏表示,之后采用时间池的操作形成视频块的描述符。采用统

计 BoWs 直方图模型表示行为并结合 NNC 留一交叉验证法(LOOCV)识别视频人体行为,实验结果的混淆矩阵如图 5 所示。

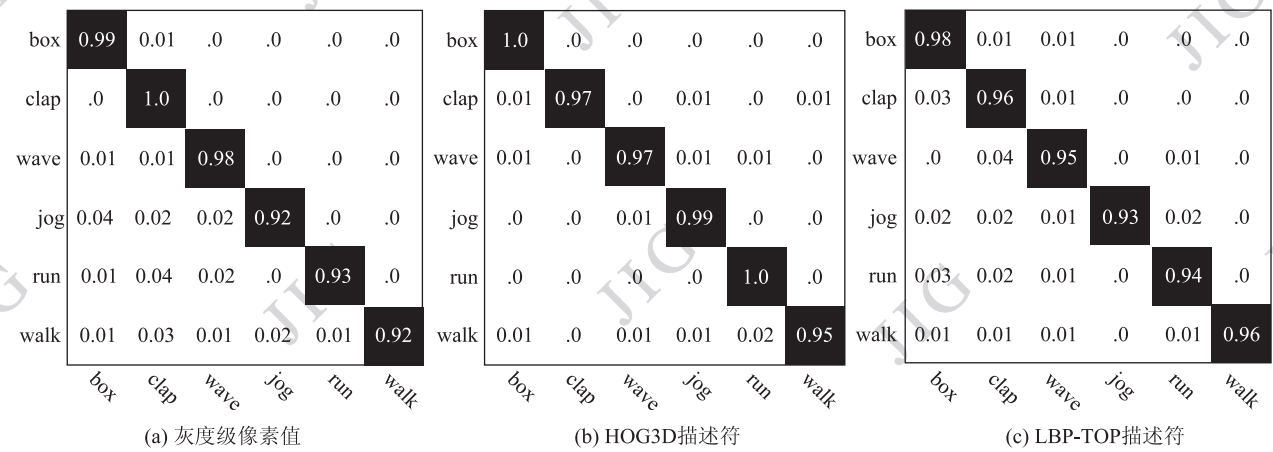


图 5 识别结果以混淆矩阵形式显示
Fig. 5 Recognition results show in confusion matrix form

在图 5 中,混淆矩阵的每行表示某种行为的正确识别率和错误识别率。从图 5 可以看出,采用图像块序列的稀疏表示并结合小尺度的时间池(本文采用连续 7 帧)获取的视频块的表示能够作为局部时空特征的描述符,与基于空时描述符 HOG3D、LBP-TOP 的稀疏表示的识别均获的了较好的识别精度。图 5(a)(c)与图 5(b)比较,上肢行为(box、clap、wave)与上下肢行为(jog、walk、run)的区分较差,分析原因是由 2 方面引起:1) HOG3D 描述符的时间尺度大于 LBP-TOP 和视频块;2) HOG3D 描述符产生与 LBP-TOP、视频块是不同的, HOG3D 描述符是以采样中心,多个子空间时间位置描述符的串联,它组合了多个空间、时间子块描述符的邻域关联信息,引入了局部形状及运动信息。

从稀疏编码局部特征的角度比较识别效果,表 1 列出了本文方法及相关文献方法在 KTH 人体行为数据上的平均识别精度。本文方法利用学习的分层树状结构字典分层编码视频块、HOG3D、LBP-TOP 描述符并结合 BoWs 模型来表示视频,在 KTH 行为数据库上分别达到 95.64%、97.99%、95.24% 的识别精度。文献[15]利用标准的稀疏编码方式将兴趣点的位移描述符转换为稀疏表示。但是,标准稀疏编码方法不能获得的鲁棒性和判别性的特征描述符。文献[16]采用局部约束稀疏编码特征与视觉词之间的距离作为上下文特征描述符。文献[7]通过迁移学习构造结构化字典编码 HOG3D。与本文

的结构化编码 3D-Patch、LBP-TOP 的相比,具有大致相当的识别效果,而本文基于 HOG3D 稀疏表示的行为识别达到了最好的效果。

表 1 KTH 数据库实验结果与已有的算法比较
Table 1 Comparisons to previous algorithm results on KTH dataset

方法	描述符	年度	平均识别精度/%
文献[7]	HOG3D	2010	94.92
文献[8]	3D-Patch	2011	93.9
文献[15]	Displacement	2012	92.59
文献[16]	HOG-HOF	2012	95.06
本文	3D-Patch		95.64
	LBP-TOP		95.24
	HOG3D		97.99

4 结 论

在 L_1 范数规则化的基础上,把规划编码路径和选择原子约束嵌入 L_1 范数规则项,提出了一种分层树状结构字典学习算法。将 L_0 范数贪婪算法的 Rank-1 编码或 Rank-n 的组编码扩展为分层树结构框架,层与层之间传递残差过程,层内采用 L_1 范数并结合嵌入的约束凸优化稀疏编码。分层学习的方法

法降低了训练字典的复杂度,学习的字典原子较少,信号的稀疏表示具有较好的鲁棒性和判别性。基于学习结构字典编码视频切片的局部特征,采用统计模型 BoWs 的直方图表示视频行为,在 KTH 数据库上,与同类文献算法相比获得了较好的实验结果,其中 HOG3D 编码的识别精度达到 97.99%。下一步,将研究学习字典、平衡参数的大小与识别性能的关系,同时将字典学习算法的应用推广到复杂视频的行为识别领域。

参考文献 (References)

- [1] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Advances in Neural Information Processing Systems. United States: The MIT Press, 2006:801-808.
- [2] Guo J, Li X. Structured dictionary learning based on group sparsity[J]. Journal of Image and Graphics, 2012, 17(11):1347-1352. [郭景峰,李贤. 基于群稀疏的结构化字典学习[J]. 中国图象图形学报, 2012, 17(11):1347-1352.] [DOI: 10.11834/jig.20121102]
- [3] Li S, Yin H, Fang L. Group-sparse representation with dictionary learning for medical image denoising and fusion[J]. IEEE Transactions on Biomedical Engineering, 2012, 59(12):3450-3459.
- [4] Gao S, Tsang I, Chia L. Laplacian sparse coding, hypergraph laplacian sparse coding, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):92-104.
- [5] Jenatton R, Mairal J, Bach F R, et al. Proximal methods for sparse hierarchical dictionary learning[C]//Proceedings of the 27th international Conference on Machine Learning. Haifa, Israel: ACM, 2010:487-494.
- [6] Wang J, Zhang H. Structured Dictionary Learning Based on Composite Absolute Penalties[M]//Informatics and Management Science III. London: Springer-Verlag, 2013:367-374.
- [7] Zhu Y, Zhao X, Fu Y, et al. Sparse coding on local spatial-temporal volumes for human action recognition[C]//Proceedings of the 10th Asian conference on Computer vision. Berlin: Springer-Verlag, 2011:660-671.
- [8] Le Q V, Zou W Y, Yeung S Y, et al. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. United States: IEEE, 2011:3361-3368.
- [9] Wei C, Chao Y, Yeh Y, et al. Locality-sensitive dictionary learning for sparse representation based classification[J]. Pattern Recognition, 2012:1277-1287.
- [10] Zelnik-Manor L, Rosenblum K, Eldar Y. Dictionary optimization for block-sparse representations[J]. IEEE Transactions on Signal Processing, 2012, 60(5):2386-2395.
- [11] 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.
- [12] Ng A Y. Feature selection, L1 vs. L2 regularization, and rotational invariance[C]//Proceedings of the twenty-first international conference on Machine learning. New York, NY, United States: ACM, 2004:78-86.
- [13] Cheng X, Li N J, Zhang S F, et al. Hierarchical associative multi-target detection and tracking with particle swarm optimization[J]. Journal of Image and Graphics, 2013, 18(9):1099-1106. [程旭,李拟珩,张索非,等. 等级关联的粒子群迭代多目标检测跟踪[J]. 中国图象图形学报, 2013, 18(9):1099-1106.] [10.11834/jig.20130907]
- [14] Bregonzio M, Gong S, Xiang T. Recognizing action as clouds of space-time interest points[C]//Computer Vision and Pattern Recognition. United States: IEEE, 2009:1948-1955.
- [15] Zhang X, Zhang H, Cao X. Action recognition based on spatial-temporal pyramid sparse coding[C]//Proceedings of IEEE Conference on Pattern Recognition. Tsukuba Science City, Japan: IEEE, 2012:1455-1458.
- [16] Zhang Z, Wang C, Xiao B, et al. Action recognition using context-constrained linear coding[J]. IEEE Signal Processing Letters, 2012, 19(7):439-442.
- [17] Zhao G, Ahonen T, Matas J, et al. Rotation-invariant image and video description with local binary pattern features[J]. IEEE Transactions on Image Processing, 2012, 21(4):1465-1477.
- [18] Thornton J, Main L, Srbic A. Fixed Frame Temporal Pooling[M]//AI 2012: Advances in Artificial Intelligence. Berlin: Springer-Verlag, 2012:707-718.