

基于特征筛选的码本区分性增强方法*

徐涛, 庹红娅, 方正, 刘力, 敬忠良

(上海交通大学 航空航天学院, 上海 200240)

摘要: 针对 BOF 模型中的码本训练问题, 提出了一种改进的 K-means 方法。传统的 K-means 方法没有考虑对采集到的特征进行筛选, 基于优化的方法可以看做是一种特征筛选的方法, 但是实现复杂, 计算量大。提出了一种基于 Gist 信息的特征筛选方法。根据 Gist 信息可以将图像粗分为背景区域和前景区域, 然后对前景区域进行密集的特征采样, 对背景区域进行稀疏的特征采样, 最后所获得的特征都用来建立码本。实验结果表明, 该方法训练的码本在 Caltech101 上有很好的分类效果, 表明了该方法的有效性。

关键词: 图像分类; BOF; 空间金字塔匹配; Gist 特征; K-均值聚类; 码本

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2014)05-1597-04

doi:10.3969/j.issn.1001-3695.2014.05.077

Improving discriminability of dictionary by features selecting

XU Tao, TUO Hong-ya, FANG Zheng, LIU Li, JING Zhong-liang

(School of Aeronautics & Astronautics, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: For the dictionary learning problem of bag-of-features(BOF) model, this paper presented an improved K-means method. K-means cannot achieve a good performance for it disregard feature selection and optimized-based method can be regard as a feature selection method, but the computation cost is usually intractable leading to long training time. This paper presented a method to select feature based on Gist information. First, images could be roughly partitioned into foreground and background based on Gist information. Then densely feature extracting was taken on the foreground and sparsely feature extracting was applied on the background. Finally, the method used these features to construct a more discriminative dictionary. Experiments on Caltech101 show that this method can achieve a better performance, and it shows that this method can achieve a better discriminability.

Key words: image classification; BOF; spatial pyramid match(SPM); Gist feature; K-means clustering; dictionary

0 引言

BOF(bag-of-features)模型首先应用在文本检索领域, 由于其快速、高效性以及很好的分类效果, 因此有学者^[1-3]将其引入图像分类领域。BOF 模型简单, 分类效果不受目标在图像上的位置影响, 且图像表达的长度与图像上的特征数无关。其缺点是没有考虑特征的空间信息, 因此也就无法定位目标, 而这一点对分类效果的提升至关重要。

BOF 模型主要包括特征提取、码本训练、特征编码和特征合并四个步骤。为了改进 BOF 模型, 研究人员提出了很多方法。文献[3]中提出了一种结合部件模型的方法; 文献[4, 5]中提出了基于几何对应的方法。这两种方法都取得了较好的分类效果, 但是计算较复杂。受金字塔匹配核方法^[6]的启发, Lazebnik 等人在文献[7]中提出了空间金字塔匹配(SPM)方法, 该方法巧妙地将特征的空间信息加入模型之中, 分类效果比 BOF 方法有很大的提高而且计算简单。随后很多学者基于 SPM 框架, 对 BOF 模型的训练和编码等阶段提出了许多改进方法。为了得到更好的分类效果, Yang 等人^[8]将稀疏编码方法引入 SPM 模型中, 得到了基于稀疏编码的空间金字塔匹配(ScSPM)方法, 使得码本的训练与编码很好地结合起来。受稀

疏编码局部性的启发, Wang 等人^[9]提出局部约束编码(LLC)方法, 进一步提高了分类效果, 尤其适用于较大的数据集, 如 Caltech256。Liu 等人^[10]提出局部软编码(LSC)方法, 这些都属于编码方式的改进方法。文献[11]中提出一种弱监督定位的方法, 把目标从背景中分离出来, 减少背景信息的干扰, 对分类效果有一定的提高。但它是一种类似于启发式的方法, 因此它的计算量很大, 而且对类内差异比较大的类别定位效果也不好。

本文的工作集中在对码本训练方法的改进。码本作为编码的基准, 码本的好坏关系到图像分类的效果。常用的码本训练方法是 K-means 聚类的方法, 这种方法简单易实现, 但其缺点是生成的码本没有考虑特征样本的区分性信息。而 ScSPM、LLC 等优化训练方法其分类效果的体现依赖于与编码方法的配合, 因此生成的码本都不具有推广性。Mairal 等人^[12]提出了一种独立于编码方法的优化码本训练方法, 生成具有类间区分性的码本, 获得很好的分类效果, 但其计算复杂度和可理解性都不如传统的 K-means 聚类方法。

针对以上问题, 本文提出一种基于 Gist 特征信息检测、增强码本区分性的码本训练方法。通过对图像局部区域信息量的检测, 对目标和背景样本进行非均匀密集采样, 从而增加具

收稿日期: 2013-06-04; 修回日期: 2013-07-23

基金项目: 航空科学基金资助项目(20115557007)

作者简介: 徐涛(1988-), 男, 江西人, 硕士研究生, 主要研究方向为图像分类与图像处理(nexutaosdu@163.com); 庹红娅(1976-), 女, 湖北人, 副研究员, 主要研究方向为模式识别与图像处理; 方正(1990-), 男, 上海人, 硕士研究生, 主要研究方向为图像分类与处理; 刘力(1989-), 男, 江苏人, 硕士, 主要研究方向为图像分类与图像处理; 敬忠良(1960-), 男, 四川人, 教授, 长江学者, 主要研究方向为机动目标跟踪和图像融合。

有类别区分性的特征样本数,减少不具有类别区分性的特征样本数,获得具有类别区分性的样本特征来进行码本训练。

1 本文的主要工作

Gist 特征描述子由 Oliva 等人^[13]提出,它是表征图像的粗糙度等结构信息的综合。本文并不对这些信息进行区分,只是考虑图像局部区域信息量的大小。由于目标多存在信息量丰富的区域,因此对于背景不太复杂的图像,Gist 特征可以用来对目标进行近似的定位。

本文提出一种基于 Gist 特征信息检测、增强码本区分性的码本训练方法。首先基于 SPM 框架对训练集图像进行 Gist 特征检测,进行目标和背景的划分后,使用非均匀密集采样方式提取具有类别区分性的 SIFT 样本特征来进行 K-means 聚类,构造码本;然后对训练图像进行编码,得到每幅图像的向量表达,通过这些向量训练出一个 SVM 分类器。测试集合用相同的方法得到图像的向量表达,用来测试分类方法的效果。

1.1 基于 Gist 特征的目标定位

对图像 $i(x, y)$ 进行傅里叶变换,可以获得图像的频域信息 $I(f_x, f_y)$, $\|I(f_x, f_y)\|$ 也被称为振幅谱,由式(1)得到。

$$I(f_x, f_y) = \sum_{x, y=0}^{N-1} i(x, y) h_r(x, y) e^{-j2\pi(f_x x + f_y y)} \quad (1)$$

其中: f_x 和 f_y 是空间频域变量, $h_r(x, y)$ 是圆形 Hanning 窗口,用来降低边界效应。

设 $\|I(x', y', f_x, f_y)\|$ 为图像空间 (x, y) 局部区域的振幅谱,可以通过加窗的傅里叶变换描述得到。

$$I(x', y', f_x, f_y) = \sum_{x', y'=0}^{N-1} i(x, y) h_r(x - x', y - y') e^{-j2\pi(f_x x + f_y y)} \quad (2)$$

其中: x, y 表示图像空间的局部位置坐标。

$\|I(x', y', f_x, f_y)\|^2$ 是图像局部区域的能量谱,表征图像局部结构信息。对 $\|I(x', y', f_x, f_y)\|^2$ 进行卡洛(KL)变换,可分解为一组正交函数的组合:

$$\|I(x', y', f_x, f_y)\|^2 \cong \sum_{i=1}^N w_i \Psi_i(x, y, f_x, f_y) \quad (3)$$

其中: w_i 为权值参数; $\Psi_i(x, y, f_x, f_y)$ 是 KL 变换的基函数,其中 $\langle \Psi_i, \Psi_j \rangle = 0 (i \neq j)$; N 表示基函数的个数 ($N < 50$), N 的大小决定了 Gist 特征的维数。

权值参数 w_i 可以通过下式获得:

$$w_i = \sum_x \sum_y \iint A(x, y, f_x, f_y)^2 \times \Psi_i(x, y, f_x, f_y) df_x df_y \quad (4)$$

设 $G = \{w_1, w_2, \dots, w_N\}$, 用 G 表示局部图像区域的 Gist 特征。

图 1 显示了图像的 Gist 特征。为了结合空间金字塔匹配(SPM)模型,将图像均等划分为图像子区域,然后分别计算每一个局部图像块中的 Gist 特征。图 1(a) 行代表原始图像, (b) 为对应图像各个子区域信息量的可视化显示,不同的颜色代表不同信息(见电子版),亮度越高就表示所包含的信息越丰富。这里不区分这些信息的种类及其代表的物理意义,关心的是各种信息的和,以便能够比较不同区域的信息量大小。

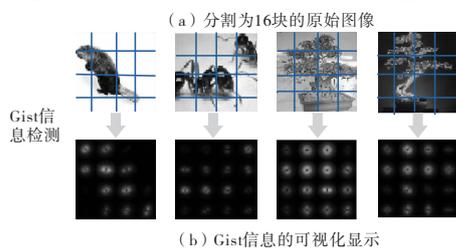


图 1 基于 Gist 特征的目标近似定位

图 2 表达了 Gist 特征信息检测的流程。图 2(a) 是原始图像, (b) 是每一个子图像区域 Gist 特征可视化显示的结果, (c) 是图(b) Gist 特征信息量的归一化直方图分布。从图 2(c) 中可看出,第 7 个 bin 直方图数值最高,其对应图 2(b) 中第二行第 3 个子区域,对应图 2(a) 中目标信息量最大的子区域。第 6、10 个 bin 直方图数值次高,分别对应图 2(b) 中第二行第 2 个子区域和第三行第 2 个子区域。从图 2(a) 中可看出,这两个子区域是目标信息量较为丰富的子区域。这三个区域都是目标所占据的子区域。图 2(c) 中第 1、2、4、15 等 bin 的直方图数值较小,分别对应图 2(b) 中第一行第 1、2、4 的子区域和第四行第 3 个子区域,这些区域都是背景所占据的子区域。

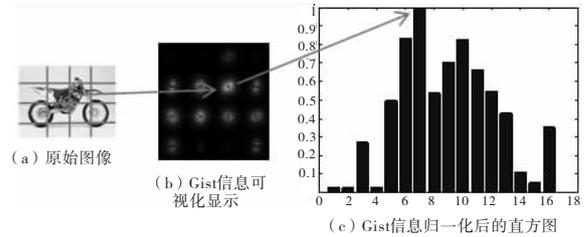


图 2 Gist 特征信息检测示意图

由于 Gist 亮度高的区域基本对应着目标所在的区域, Gist 亮度低的区域多对应背景区域,所以可以通过 Gist 特征进行图像信息量检测,获得目标所占据的主要子区域,对目标进行近似的定位。

1.2 非均匀密集采样方式

特征提取是图像分类识别的基础步骤。常用的局部特征是 Lowe^[14] 提出的 SIFT(scale-invariant feature transform, 尺度不变特征转换)特征。SIFT 特征与图像尺度和旋转无关,并且对于光线、噪声和视角的变化也具有很好的鲁棒性,因此广泛用于图像分类领域,取得了很好的效果。本文也采用 SIFT 特征描述子。

SIFT 特征提取的方式分为基于兴趣点的采样和密集采样这两种方式。Li 等人^[15]通过实验证明,对于图像分类来说,密集采样得到的分类效果要优于基于兴趣点采样的方式。因为基于兴趣点的采样方式采集的特征并不能够很好地代表图像中的目标,很有可能大部分的兴趣点都落在背景上,这样反而会使分类效果恶化。而密集采样的方式是对图像进行均匀采样,就不存在以上的问题。因此本文采用的是密集采样的方式。但是密集采样存在的问题是,随着对目标特征的采样也会对背景进行采样,对于背景占据篇幅比较大的图像来说,这个问题尤其严重,导致采集大量不具有类别区分性的特征,对后续进行 K-means 聚类生成码本带来不好的影响,码本区分性变差。

为了清楚地描述这种影响,首先分析 K-means 聚类的过程。如上文所提到的 K-means 聚类方法趋向于将聚类中心聚集在样本密度高的区域,而样本点集密度低的区域分配的聚类中心数少。如果不改进密集采样方式,那么与类别无关的特征就会占据大量的聚类中心,使得与类别相关的描述子反而占据少量的聚类中心,导致 K-means 聚类的码本不能体现类别区分性质。

为了更好地描述本文方法对 K-means 聚类效果产生的影响,本文在二维空间对三个样本集群进行了实验,如图 3 所示。图 3(a) 是直接对密集采样得到的特征进行聚类的结果, (b) 是采用 Gist 特征检测并对特征进行选择后的聚类结果。将所有样本分为三个大的样本集群,这三个大的样本集群占据了绝

大多数的聚类中心,其中 A 表示与类别无关信息的样本集群, B 和 C 表示与类别相关的样本集群。使用 Gist 特征检测后,过滤了部分背景特征样本,使得 A 部分的样本数减少,这样 A 集群所占据的聚类中心数目减少,而 B 和 C 集群所占据的聚类中心数目增多。码本中与类别无关的码字数目减少,码本的类别区分特性增强,从而提高了分类效果。

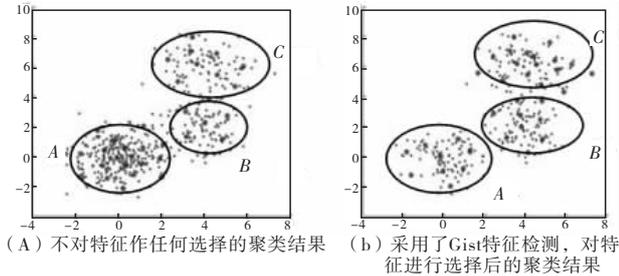


图 3 二维仿真数据集上 K-means 聚类的结果

1.3 分类方法和步骤

本文采用基于 SPM 框架的分类方法。设 I 为数据库图像集, $I_r = \{p_i, p_i = \{p_{ij}\}, i = 1, 2, \dots, N, j = 1, 2, \dots, n\}$ 为训练图像集, N 为类别数, n 为第 i 类图像训练样本个数, n 的大小可以根据实验需求调节。数据库 I 中剩余的图像 I_e 作为分类测试集,用来测试分类效果。其算法步骤如下:

a) 在训练阶段,先将训练集图像 p_{ij} 划分为 4×4 个子区域,对每个子区域进行 Gist 信息检测,获得 p_{ij} 的 Gist 特征向量 G_{ij} ,对 G_{ij} 的元素进行排序,前 M 个信息量较大的子区域采用步长为 4 的密集 SIFT 特征采样,剩下的子区域进行步长为 8 的均匀 SIFT 特征采样。

b) 对训练集 I_r 中所有图像 p_{ij} 重复步骤 a), 获得 SIFT 特征样本集 $F = \{s_i, s_i = \{s_{ij}\}, i = 1, 2, \dots, N, j = 1, 2, \dots, n\}$ 。

c) 设聚类中心为 m 个,采用 K-means 方法对 SIFT 特征样本集 F 进行聚类,获得码本 $D \in \mathbb{R}^{128 \times m}$ 。

d) 根据码本 D ,利用局部软编码方法(LSC)对训练集 I_r 中所有图像 p_{ij} 进行编码。基于 SPM 的框架,使用 max-pooling 合并和向量串联的方法得到 p_{ij} 的图像表达 $f_{ij} \in \mathbb{R}^{21 \times m}$ 。

e) 将 d) 中获得的图像表达 $\{f_{ij}\}$ 作为 SVM 分类器的训练样本,训练得到图像分类器。

f) 测试阶段,对测试集 I_e 中每一幅图像重复步骤 a) b) d), 获得测试集每一幅图像的向量表达,利用步骤 e) 得到的分类器进行分类,测试本文分类方法的分类效果。

2 实验结果与分析

本章中将在 Caltech101 数据库上验证分类效果。数据库中有些图像是彩色的,都将其转换为大小不超过 300×300 的黑白图像。实验中采用尺度为 16×16 的 SIFT 特征。每个图像划分为 4×4 个子区域,对 Gist 信息量大的前 8 个子区域采取步长为 4 的密集采样方法,剩下的子区域采取步长为 8 的均匀采样方法,这样每幅图像提取的特征数量和均匀步长为 6 时^[10]的特征数目一致,但是更多地获得了来自目标区域的特征而减少来自背景区域的特征。

编码过程采用局部软编码方法(LSC)。LSC 是基于 SPM 模型的方法,其编码方法与码本生成方法是相互独立的,所以采用 LSC 进行目标分类可以验证本文训练的码本是否会提高分类效果。另外本文还将实验结果与 ScSPM 和 LLC 方法

进行了比较。为了保持实验的一致性,便于与他人的实验结果进行比较,本文实验都将采用 SVM^[16] 分类器。

2.1 Caltech101 数据库实验结果

Caltech101 数据库有 9 144 幅图像包含 101 种类别,如动物、植物和花草等种类,类内的差别性比较大。每个种类所包含的图像从 31 幅到 800 幅不等。采用文献[9,10]中的建议,每类取出 15 幅和 30 幅图像作为训练样本,剩余的作为测试样本。训练的码本大小设为 1 024。采用三层的 SPM 模型,从第 0 层到第 2 层每层分割为 $2^l \times 2^l (l=0,1,2)$ 。

分类实验结果如表 1 所示。当训练样本数为 15 时,SPM 方法效果最差,为 56.4%,效果最好的是 ScSPM 和本文方法,取得了 67% 的识别率。当训练样本数为 30 时,本文的方法分类效果最好,达到 75%。需要注意的是,有时无法复现文献[10]中的实验效果,原因可能是在实验过程中有细微的差异,因此本文方法比 LSC 方法提高了 0.8% ~ 1.2%,比 LLC 和 ScSPM 方法的分辨率也好。分类结果表明,根据区域信息量不均匀采集特征的方法,降低了无关特征对码本生成方法的干扰,增强了码本的区分性,提高了分类效果。

表 1 数据库 Caltech101 上的分类结果

方法	样本数为 15	样本数为 30
SPM ^[5]	56.40	64.60 ± 0.80
ScSPM ^[8]	67.0 ± 0.45	73.20 ± 0.54
LLC ^[9]	65.43	73.44
LSC ^[10]	--	72.5 ± 0.81
LSC-Ours	67.0 ± 0.76	73.74 ± 0.92

2.2 实验分析

为了更好地理解本文的方法,研究根据 Gist 信息量进行非均匀采样对 Caltech101 数据库分类效果的影响。根据 Gist 特征向量分别选取前 6、8 和 10 个图像子区域进行高密度采样,剩下的区域采用低密集均匀采样,分类效果如图 4 所示。图 4 中纵轴代表图像的平均分类准确率,横轴代表每类训练样本的个数,本文方法的训练样本数分别取 5、10、15、20、25、30。从图 4 中可以看出,当选择高密度采样区域为 8 时,其分类效果比高密度采样区域为 6 时的分类效果好很多,而当选择高密度采样区域为 10 时,分类效果只有微弱的提升,这可能是混入的背景特征抵消掉了部分分类效果。所以高密度采样的子区域并不是越多越好,如果过多,就会有背景区域被密集采样,得到更多的与类别无关的特征。但是高密度采样的区域数目也不能过少,因为过少就会有目标区域被低密集采样,遗漏了很多与目标相关的特征。另外也可以发现,随着训练样本个数不断增加,分类效果越来越好。当选择密集采样区域为 10、训练样本数为 30 时,本文的方法取得最好的效果为 74.01%,比选择密集采样区域为 8 时的效果提升了 0.27%。

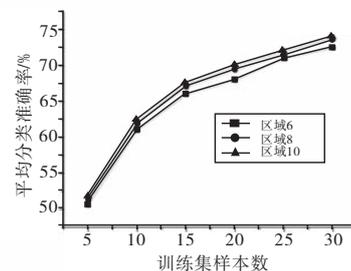


图 4 非均匀密集采样对分类效果的影响

3 结束语

本文基于 SPM 框架提出了一种基于 Gist 特征信息检测的码本区分性增强方法。通过对图像 Gist 特征信息量的检测,对目标和背景进行划分,使用非均匀密集采样有效提取目标特征,过滤与类别无关特征,从而使得 K-means 聚类生成的码本更加具有类别区分性,提高了分类效果。实验结果表明,本文方法进行样本筛选得到的码本在类别区分性上优于没有经过筛选的码本,分类效果优于 LSC、ScSPM、KSPM 等方法。本文方法和 LLC 分类效果基本一致,但是本文方法实现更加简单、快速、易于理解,码本生成方法与后续编码方法完全独立,因此生成的码本更具有通用性。在将来的工作中,笔者将对信息检测的方法进行改进,尝试用其他更加简洁、更准确的方法来检测信息,降低检测所需要的时间,同时更加精确地获得与类别相关的特征,这样将进一步提高分类效率和准确率。

参考文献:

- [1] LI Fei-fei, FERGUS R, PERONA P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2004:178-186.
- [2] NOWAK E, JURIE F, TRIGGS B. Sampling strategies for bag-of-features image classification [C]//Proc of European Conference on Computer Vision. 2006:490-503.
- [3] FERGUS R, PERONA P, ZISSERMAN A. Object class recognition by unsupervised scale-invariant learning [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2003:264-271.
- [4] BERG A, BERG T, MALIK J. Shape match and object recognition using low distortion correspondences [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2005:509-522.
- [5] LAZEBNIK S, SCHMID C, PONCE A J. A maximum entropy framework for part-based texture and object recognition [C]//Proc of International Conference on Computer Vision. 2005:832-838.
- [6] GRAUMAN K, DARRELL T. Pyramid match kernels: discriminative classification with sets of image features [C]//Proc of International Conference on Computer Vision. 2005:1458-1465.
- [7] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2006:2169-2178.
- [8] YANG Jian-chao, YU Kai, GONG Yi-hong, *et al.* Linear spatial pyramid matching using sparse coding for image classification [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2009:1794-1801.
- [9] WANG Jin-jun, YANG Jian-chao, YU Kai, *et al.* Locality-constrained linear coding for image classification [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2010:3360-3367.
- [10] LIU Ling-qiao, WANG Lei, LIU Xin-wang. In defense of soft-assignment coding [C]//Proc of International Conference on Computer Vision. 2011:2486-2493.
- [11] RUSSAKOVSKY O, LIN Yuan-qing, YU Kai, *et al.* Object-centric spatial pooling for image classification [C]//Proc of European Conference on Computer Vision. 2012:1-15.
- [12] MAIRAL J, BACH F, PONCE J, *et al.* Supervised dictionary learning [C]//Advances in Neural Information Processing System. 2008:1033-1040.
- [13] OLIVA A, TORRALBA A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. *International Journal of Computer Vision*, 2001, 42(3):145-175.
- [14] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2):91-110.
- [15] LI Fei-fei, ERONA P P. A Bayesian hierarchical model for learning natural scene categories [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2005:524-531.
- [16] FAN R E, CHANG K W, HSIEH C J, *et al.* LIBLINEAR: a library for large linear classification [J]. *Journal of Machine Learning Research*, 2008, 9(1):1871-1874.
- [17] JURIE F, TRIGGS B. Creating efficient codebooks for visual recognition [C]//Proc of European Conference on Computer Vision. 2005:604-610.
- [18] YANG Jian-chao, YU Kai, GONG Yi-hang, *et al.* Linear spatial pyramid matching using sparse coding for image classification [C]//Proc of International Conference on Computer Vision and Pattern Recognition. 2009:1794-1801.
- [19] HOLUB G G, PERONA P A D. Caltech-256 object category dataset, TR7694 [R]. [S. l.]: California Institute of Technology, 2007.
- [20] ANKERST M, BERCHTOLD S, KEIM D A. Similarity clustering of dimensions for an enhanced visualization of multidimensional data [C]//Proc of IEEE Symposium on Information Visualization. Washington: IEEE Computer Society, 1998:52-60.
- [21] ARTERO A O, de OLIVEIRA M C F, LEVKOWITZ H. Enhanced high dimensional data visualization through dimension reduction and attribute arrangement [C]//Proc of the 10th International Conference on Information Visualization. 2006:5-7.
- [22] 孙扬, 唐九阳, 汤大权, 等. 改进的多变元数据可视化方法 [J]. *软件学报*, 2010, 21(6):1462-1472.
- [23] SHI Cong-lei, CUI Wei-wei, LIU Shi-xia, *et al.* RankExplorer: visualization of ranking changes in large time series data [J]. *IEEE Trans on Visualization and Computer Graphics*, 2012, 18(12):2669-2678.
- [24] LOUDCHER S, BOUSSAID O. OLAP on complex data visualization operator based on correspondence analysis [C]//Lecture Notes in Business Information Processing. Berlin: Springer, 2012:172-185.
- [25] LIVINGSTON M A, DECKE J W, AI Zhu-ming. Evaluating multivariate visualizations on time-varying data [C]//Proc of SPIE 8654 on Visualization and Data Analysis. 2013.
- [26] KEHRER J, HAUSER H. Visualization and visual analysis of multifaceted scientific data: a survey [J]. *IEEE Trans on Visualization and Computer Graphics*, 2013, 19(3):495-513.
- [27] [http://eagereyes.org/parallel-sets\[EB/OL\].](http://eagereyes.org/parallel-sets[EB/OL].)

(上接第 1596 页)