

基于可变性分析的紧致图像表达

赵鑫^{1,2}, 黄凯奇², 谭铁牛²

(1. 中国科学技术大学自动化系, 安徽合肥 230037; 2. 中国科学院自动化研究所智能感知与计算研究中心, 北京 100190)

摘要: 图像表达是图像分类中最基本也是最重要的一个环节, 当前的图像表达方法为了获得较高的分类性能, 通常采用维度极高的特征向量, 这给分类器的训练和特征的存储带来了极大的负担。同时, 这些方法没有考虑图像的变化给图像表达所带来的影响, 为此, 针对以上的问题提出了一种对图像的可变性进行建模的方法。该方法首先使用高斯混合模型对底层视觉特征进行建模; 再构造图像的充分统计量; 最后采用可变性分析对充分统计量进行分解, 并结合偏最小二乘回归方法获得紧致的图像表达。在公开的主流图像分类数据库上, 该方法在获得更高的分类性能的同时极大地降低了分类器的训练和特征存储的开销。

关键词: 图像表达; 图像分类; 可变性分析; 因子分析; 偏最小二乘

中图分类号: TP391.41 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2014.02.008

引用格式: Zhao Xin, Huang Kaiqi, Tan Tieniu. Compact image representation based on variability analysis[J]. Journal of University of Science and Technology of China, 2014, 44(2): 128-137.

赵鑫, 黄凯奇, 谭铁牛. 基于可变性分析的紧致图像表达[J]. 中国科学技术大学学报, 2014, 44(2): 128-137.

Compact image representation based on variability analysis

ZHAO Xin^{1,2}, HUANG Kaiqi², TAN Tieniu²

(1. Department of Automation, University of Science and Technology of China, Hefei 230027, China;

2. Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Image representation is the most fundamental and important aspect in image classification tasks. Most existing image representation methods use quite high dimensional feature vectors for image representation in order to achieve desired performance, which results in an inevitable drawback which is a classification problem with very high-dimensional feature vectors. Meanwhile, the existing methods have not considered image variations in image representation. Thus, an image representation method was proposed to model the variability in image classification. First, a Gaussian mixture model (GMM) was used to model the low-level visual feature vectors. Then, the sufficient statistics of images were constructed. Finally, the proposed variability analysis was utilized to decompose the sufficient statistics, and a compact image representation was obtained by means of partial least square regression. The proposed

收稿日期: 2013-04-16; 修回日期: 2013-05-20

基金项目: 国家重点基础研究发展(973)计划(2012CB316302), 国家自然科学基金(61135002, 61175007)资助。

作者简介: 赵鑫, 男, 1984年生, 博士生。研究方向: 数字图像理解与分析, 计算机视觉。E-mail: xzhao@nlpr.ia.ac.cn

通讯作者: 谭铁牛, 博士/研究员。E-mail: tnt@nlpr.ia.ac.cn

method not only achieves better performance on the public image classification datasets, but also reduces the burdens of classifier training and feature storage.

Key words: image representation; image classification; variability analysis; factor analysis; partial least square

0 引言

众所周知,数字图像在计算机中是以矩阵或者维度较高的向量进行存储.如果直接对其进行分类等高层操作,数据量就会极大.同时数字图像本身包含较多的冗余信息,直接处理通常效果不佳.因此,在与图像分类相关的研究问题中,对图像表达的研究是最基本和最重要的研究方向,受到众多研究者的持续关注^[1-7].

在图像表达早期的工作中,基于密集采样的底层视觉特征提取的方式^[8-10]无疑是这其中最具代表性的方法.这种方式的优点是不需要对图像进行复杂的分割处理,从而使得提取工作变得高效并且对复杂的图像背景较为鲁棒.由于提取到的底层特征个数在不同尺寸的图像中不一样,因此就需要有一种编码的方法将这些变长的特征集合组合成定长的整体图像表达.这种方法通常需要借助统计的手段,比如常见的 BoW 模型^[1].因此,BoW 模型既是一种底层特征编码技术,也是一种图像表达的方法.

基于 BoW 模型发展出了两类图像表达方法:

(I) 基于重构误差最小化的方法:这类方法主要关注的是底层视觉特征在特征空间中的分布特点.以重构误差最小化为目标,这类方法使用聚类得到的视觉词包去重构每一个底层视觉特征.最常用的方法叫做矢量量化^[1],该方法使用线性回归模型进行底层视觉特征重构.此后,在以上线性回归模型的基础上,一些方法在特征空间中引入了不同的约束,进而形成了不同的方法.比如,稀疏性^[11]、显著性^[6]和局部性^[5]等.为了进一步减小重构误差,文献[4]中的方法使用分段线性回归的方法取代了矢量量化中的线性回归方法.

(II) 基于生成式模型的方法:当前图像分类任务多采用判别式的线性分类器进行分类.在文献[2]中,判别式的分类器与生成式核相结合,该生成式核将对生成式模型的参数的倒数串接在一起.为了充分利用判别式模型和生成式模型的能力,底层视觉特征可以通过生成式核的解析直接映射的方式进行构造.在这方面的的工作中,文献[3]已经在大多数的

公开数据库上取得了领先的性能.

以上基于 BoW 模型的图像表达方法存在以下两个局限性:

(I) 根据 BoW 模型发展出来的一系列图像表达框架从不同角度对图像分类中存在的问题进行了分析.在实际的图像分类问题中,真正影响图像表达最终效果的因素是多种多样的.比如,图像类内的差异大、类间的差异小以及复杂的图像背景等.传统的图像表达模型中都没有充分考虑这些因素,而这些因素是潜在的和不可观测的,这就给图像分类带来了很多的不确定性.在此假设有一系列的潜在因子决定了这些影响图像分类性能的潜在因素.这些潜在因子会因为问题的不同而发生不同的变化,这一现象就是模式表达中最具挑战性的模式可变性问题.

(II) 当前图像分类任务从分类代价的角度出发,通常都使用构造简单的线性分类器进行分类.而低维度的图像表达特征向量显然不具备足够的判别性,特别是当类别较多的时候.因此就需要在效率与性能之间寻找到一个平衡点.为了能够获得更高的分类性能,当前的图像表达方法通常都构造了高维度的图像表达特征向量.特别是最近提出的基于生成式模型的方法^[2-4],这类方法在各公开数据库上都展现了优异的性能.而基于重构误差最小化的方法通常会采用非常大规模的视觉词包获得高维度的表达^[5-6],因此,即使是在小规模或者中等规模的数据库上,这些方法也会给分类器训练带来巨大的负担.

不同于一般的底层特征编码方法,本文将要研究图像表达的更加一般性的框架.该框架的研究目标是希望充分利用影响图像分类的各种潜在因素来解决图像分类中的可变性问题,并且最终获得一种紧致的图像表达.同时,该框架希望增强图像表达的判别性,进而提高分类性能.

1 框架概述

如图1所示,该框架首先在每幅图像中提取底层特征构成每幅图像的底层特征集,再用预先训练

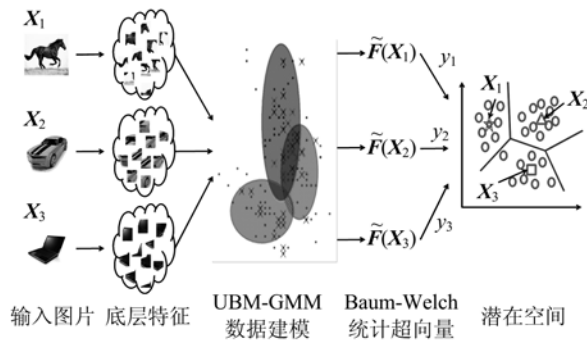


图 1 图像表达框架

Fig. 1 Image representation framework

得到的混合高斯模型 (Gaussian mixture model, GMM) 对底层特征进行数据建模, 建模之后会得到每幅图像的 Baum-Welch 统计超向量. 由于这些 Baum-Welch 统计超向量的维度较高, 并且在构造它们的过程中没有考虑影响图像分类的潜在因素, 因此可结合每幅图像的分类标签, 将 Baum-Welch 统计超向量变换到一个低维的潜在空间中, 使得变换到潜在空间中的特征向量具有更强的判别性, 并且对影响图像分类性能的诸多潜在因素进行补偿.

2 数据模型

在 BoW 模型中, 底层特征是通过硬投票的方式被划分给距离其最近的视觉关键词. 在随后的基于 BoW 模型的扩展工作中, 软投票的方式^[12-13]得到了较多的关注. 而基于 GMM 进行的表达方式将 BoW 模型扩展成一个基于概率的框架. 本文提出使用 GMM 来对图像中提取的底层特征进行数据建模, 并将基于 GMM 建模得到的 Baum-Welch 统计超向量的结果作为构造最终图像表达的基础.

在数据建模的过程中, 首先对训练集的图片提取底层特征, 用所有的训练集的底层特征训练得到通用背景模型 (universal background model, UBM), UBM 可由 GMM 训练得到. 假设 $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, N\}$ 为一幅图像中提取出的底层特征集合, \mathbf{x}_i 是第 i 个底层特征, 特征维度为 D 维. GMM-UBM 通常由 K 个高斯混合分量组成: $\mathcal{P}(\mathbf{x}) = \sum_{k=1}^K \alpha_k \mathcal{G}_k(\mathbf{x})$, 每个高斯混合分量服从高斯分布 $\mathcal{G}_k(\mathbf{x}) \sim \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. α_k 为第 k 个高斯混合分量的权重, $\boldsymbol{\mu}_k \in \mathbb{R}^D$ 和 $\boldsymbol{\Sigma}_k \in \mathbb{R}^{D \times D}$ 分别为第 k 个高斯混合分量的均值向量和对角协方差矩阵. $\boldsymbol{\sigma}_k \in \mathbb{R}^D$ 为第 k 个高

斯混合分量的对角协方差矩阵的方差向量.

GMM-UBM 超向量是基于 GMM-UBM 的模型参数而获得的, 其定义为 $\mathbf{m}_0 \in \mathbb{R}^{D \times K}$, 它是将 GMM-UBM 的所有均值向量串接起来得到的一个超向量. 该 GMM-UBM 超向量与类别无关. 对于一幅图像中提取的底层特征集合 \mathbf{X} 而言, 它的零阶、一阶和二阶 Baum-Welch 统计^[14-15]分别按照如下方式构造:

$$N_k = \sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta}) \quad (1)$$

$$\mathbf{F}_k = \frac{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta}) \mathbf{x}_i}{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta})} \quad (2)$$

$$\mathbf{S}_k = \frac{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta}) \mathbf{x}_i \mathbf{x}_i^T}{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta})} \quad (3)$$

式中, $\boldsymbol{\theta}$ 为 GMM-UBM 的参数集, $\mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta})$ 表示的是 \mathbf{x}_i 由第 k 个高斯混合分量生成的后验概率.

为了进行后续处理, 在此将式(2)和(3)中的一阶和二阶统计量中心化后重新记为

$$\mathbf{F}_k = \frac{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta}) (\mathbf{x}_i - \boldsymbol{\mu}_k)}{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta})} \quad (4)$$

$$\mathbf{S}_k = \frac{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta}) (\mathbf{x}_i - \boldsymbol{\mu}_k) (\mathbf{x}_i - \boldsymbol{\mu}_k)^T}{\sum_{i=1}^N \mathcal{P}(k | \mathbf{x}_i, \boldsymbol{\theta})} \quad (5)$$

同时, 使用 GMM-UBM 的对角协方差矩阵将式(4)的结果归一化如下:

$$\tilde{\mathbf{F}}_k = \boldsymbol{\Sigma}_k^{-\frac{1}{2}} \mathbf{F}_k \quad (6)$$

以上对 $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D, i = 1, \dots, N\}$ 的操作实际是一种白化的处理过程. 白化操作的目的是为了对底层特征的波动进行补偿, 从而使得统计量能够保持不变性. 在此定义:

$$\tilde{\mathbf{F}}(\mathbf{X}) = [\tilde{\mathbf{F}}_1^T, \dots, \tilde{\mathbf{F}}_K^T]^T \quad (7)$$

是关于 \mathbf{X} 的白化处理后的 Baum-Welch 统计超向量. $\tilde{\mathbf{F}}(\mathbf{X}) \in \mathbb{R}^{DK \times 1}$ 将 \mathbf{X} 在所有 K 个高斯混合分量上的 Baum-Welch 统计全部串接起来. 这一步结束之后得到的每幅图像的 Baum-Welch 统计超向量将作为充分统计量进行后续分析.

3 图像分类可变性潜在因子分析

如前所述,图像可变性是当前图像分类中最为关键也是最困难的一个问题.导致该问题产生的因素是多种多样的,而且这些因素无法通过直接观测进行描述.这些因素通常被称为潜在因子,下面将使用因子分析^[16]的方法对其进行剖析.

3.1 因子分析

实际的图像分类任务存在很大的可变性,变量之间的关系也错综复杂.因子分析是一种数据简化的方法,它描述数据中众多变量之间的内在关系,刻画数据的产生过程.该方法通过将原始数据变换到一个假设的潜在空间的方式,建立原始数据与潜在因子之间的相互关系,完成对数据的建模.

如图 2 所示,因子分析通常描述以下的数据生成过程:

图 2(a)中,假设随机的 D' 维潜在因子 $y \in \mathbb{R}^{D'}$ (图中实心圆点)由一个标准正态分布所生成:

$$y \sim \mathcal{N}(0, \mathbf{I}), \mathbf{I} \in \mathbb{R}^{D' \times D'} \quad (8)$$

式中, \mathbf{I} 为单位方阵.

图 2(b)中,将 y 通过因子加载矩阵进行线性变换,变换矩阵定义如下:

$$\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_D] \in \mathbb{R}^{D \times D'} \quad (9)$$

图 2(c)中,再将线性变换后的结果按坐标轴做一个尺度为 $\mu \in \mathbb{R}^D$ 的平移,从而得到 $\tilde{x} \in \mathbb{R}^D$:

$$\tilde{x} = \mathbf{W}y + \mu \quad (10)$$

图 2(d)中,对 \tilde{x} 加入扰动 Ψ ,最终可得实际的观测变量 x 由均值为 $\mathbf{W}y + \mu$, 对角协方差为 Ψ 的高斯分布生成:

$$\mathcal{G}_\theta(x | y) \sim \mathcal{N}(x; \mathbf{W}y + \mu, \Psi), \theta = \{\mu, \mathbf{W}, \Psi\} \quad (11)$$

式(11)中的参数 $\theta = \{\mu, \mathbf{W}, \Psi\}$ 需要通过最大似然估计进行求解.当知道了模型参数 θ 之后,就可以通过它们得到潜在因子 y 作为原始数据 x 在潜在空间中的表达.由于通常 $D' < D$, 所以因子分析又有数

据降维的功能,这为获得紧致的特征表达提供了有利的工具.根据因子分析描述数据生成的过程,可以首先假设影响原始数据生成的潜在因子,然后借助这些因子对原始数据进行描述.

3.2 模型构造

由于在节 2 中讨论的基于 GMM-UBM 数据模型得到的 Baum-Welch 充分统计量对诸如类别的差异较为敏感,因此下面将要研究使用因子分析的方法对 Baum-Welch 统计量进行优化.本文基于可变性分析^[15]对图像中的各种变化进行建模.在此假设有一系列的潜在因子能够影响图像分类的性能,并且所有这些潜在因子均可以在一个统一的可变性空间中进行描述.

根据节 2 中的相关定义, $m_0 \in \mathbb{R}^{D \times K}$ 是 GMM-UBM 的超向量.该超向量由 UBM 得到,因此它与类别独立.给定一幅图像, $\mathbf{X} = \{x_i \in \mathbb{R}^D, i=1, \dots, N\}$ 是该图像中提取出的所有底层特征描述子, $m(\mathbf{X})$ 为这幅图像的 GMM 超向量. $m(\mathbf{X})$ 依赖于类别,因此它包含了图像分类中存在的所有可变性.在此定义一个可变性潜在空间,假设该空间包含了图像分类中所有的可变性.在该空间中分析 $m(\mathbf{X})$, 从而完成对图像可变性的建模.对于给定的图像,其类别依赖的 GMM 超向量 $m(\mathbf{X})$ 定义如下:

$$m(\mathbf{X}) = m_0 + \mathbf{P}v(\mathbf{X}), v(\mathbf{X}) \sim \mathcal{N}(0, \mathbf{I}) \quad (12)$$

式中, $\mathbf{P} \in \mathbb{R}^{D \times D'}$ 是低秩的可变性矩阵; $v(\mathbf{X}) \in \mathbb{R}^{D'}$ 是符合标准正态分布的随机向量. $v(\mathbf{X})$ 中的各个分量对应图像分类中的潜在因子.在式(12)的模型中,假设 $m(\mathbf{X})$ 服从均值为 m_0 以及协方差矩阵为 $\mathbf{P}\mathbf{P}^T$ 的正态分布.因此,式(12)定义的模型是因子分析的一个应用,它的目的是要将容易受到各种可变性影响的类别依赖 GMM 超向量 $m(\mathbf{X})$ 投影到一个低维度 ($D' \ll D$) 的可变性潜在空间中.该空间中的样例点与类别独立.在此将式(12)的模型称作可变性模型.

3.3 模型分析

对于一幅给定的图像,它的潜在因子 $v(\mathbf{X})$ 由 Baum-Welch 统计量条件化的后验概率进行定义.该后验概率是一个高斯分布,它的均值与 $v(\mathbf{X})$ 对应^[17]. Baum-Welch 统计量的提取过程在节 2 中进行了详细的定义.根据式(4)中的定义,在此将一阶中心化的 Baum-Welch 统计量进行如下因子分析:

$$\mathbf{F}(\mathbf{X}) = \mathbf{P}v(\mathbf{X}), v(\mathbf{X}) \sim \mathcal{N}(0, \mathbf{I}) \quad (13)$$

式中, $\mathbf{F}(\mathbf{X}) = [\mathbf{F}_1^T, \dots, \mathbf{F}_k^T]^T$.

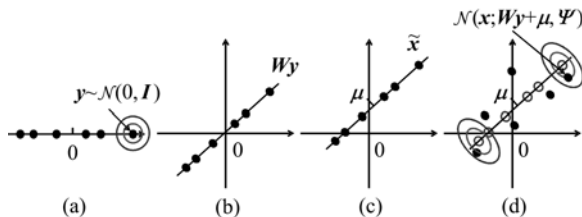


图 2 因子分析示意图

Fig. 2 Example of factor analysis

当 $F(\mathbf{X})$ 和 \mathbf{P} 已知时, $v(\mathbf{X})$ 按下面的方式计算:

$$v(\mathbf{X}) = (1 + \mathbf{P}^T \Sigma^{-1} \mathbf{N} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1} \mathbf{N} \mathbf{F} \quad (14)$$

式中, $\Sigma \in \mathbb{R}^{DK \times DK}$ 为因子分析训练过程中估计得到的分块对角协方差矩阵, 该矩阵包含了可变性矩阵 \mathbf{P} 不具备的残差可变性; $\mathbf{N} \in \mathbb{R}^{DK \times DK}$ 是分块对角协方差矩阵, 其对角线上的元素分别为 $N_k \mathbf{I} \in \mathbb{R}^{DK \times 1}$, $k = 1, \dots, K$. 图 3 是式(12)的图模型表达, Σ_0 是关于 GMM-UBM 的分块对角协方差矩阵, 其对角线上的元素分别是 $\Sigma_k \in \mathbb{R}^{D \times D}$, $k = 1, \dots, K$. M 为图像的个数.

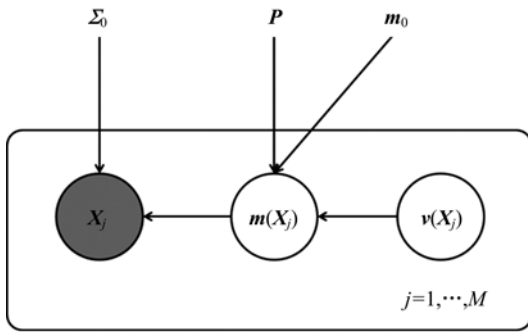


图 3 图模型

Fig. 3 Graphical model

在式(12)和式(14)中, $v(\mathbf{X})$ 是以 $\mathcal{N}(0, \mathbf{I})$ 为先验的最大后验估计的结果, 这与岭回归所获得的结果等价. 根据文献[18]的分析, 对 $v(\mathbf{X})$ 的最大似然估计定义为

$$v(\mathbf{X}) = (\mathbf{P}^T \Sigma^{-1} \mathbf{N} \mathbf{P})^{-1} \mathbf{P}^T \Sigma^{-1} \mathbf{N} \mathbf{F} \quad (15)$$

将上式写成一般的形式:

$$v(\mathbf{X}) = (\tilde{\mathbf{P}}^T \tilde{\mathbf{P}})^{-1} \tilde{\mathbf{P}} \tilde{\mathbf{F}}(\mathbf{X}) \quad (16)$$

式中, $\tilde{\mathbf{P}} = \Sigma^{-\frac{1}{2}} \mathbf{N}^{\frac{1}{2}} \mathbf{P}$; $\tilde{\mathbf{F}}(\mathbf{X})$ 为式(7)中的定义. 从式(16)可以看出, $v(\mathbf{X})$ 为标准的最小二乘结果.

根据以上分析, 将白化处理后的 Baum-Welch 统计量进行如下因子分析:

$$\left. \begin{aligned} \tilde{\mathbf{F}}(\mathbf{X}) &= \tilde{\mathbf{P}} v(\mathbf{X}) + \boldsymbol{\varepsilon}(\mathbf{X}), \\ v(\mathbf{X}) &\sim \mathcal{N}(0, \mathbf{I}), \boldsymbol{\varepsilon}(\mathbf{X}) \sim \mathcal{N}(0, \Psi) \end{aligned} \right\} \quad (17)$$

式中, $\boldsymbol{\varepsilon}(\mathbf{X})$ 为 $D \times K$ 维的随机向量; $\Psi \in \mathbb{R}^{DK \times DK}$ 为正定的分块对角矩阵. $\boldsymbol{\varepsilon}(\mathbf{X})$ 包含了可变性矩阵 $\tilde{\mathbf{P}}$ 不具备的可变性.

3.4 判别式的图像表达模型

一般而言, 图像可变性建模方式是使用因子分析法将高维的 GMM 超向量映射到低维的可变性潜在空间中. 然而, 该方法是无监督的方法, 对类别差异没有直接进行补偿. 因此, 现有工作在获得了潜

在因子 $\tilde{v}(\mathbf{X})$ 之后, 又对其进行后处理^[15]. 与这种分步的方式不同, 本文提出了基于偏最小二乘回归 (partial least square regression, PLS)^[19] 的图像可变性分析模型. 该模型利用类别标签直接在 GMM 超向量空间中操作, 在建模图像可变性的同时完成对类别差异的补偿, 最终生成完整的图像表达向量. 在此将这种表达图像的向量称作 PLS-Vector. 下面首先对偏最小二乘的原理进行简要介绍, 再对完整的 PLS-Vector 的提取过程进行详细介绍.

3.4.1 基于偏最小二乘的因子分析

传统回归方法处理病态问题(样本量少, 特征维度高的问题)存在很大的局限性. 在最近的很多工作中, 偏最小二乘针对病态问题显示了非常好的效果. 在一般的表达形式中, 偏最小二乘通过将观测变量投影到潜在空间中来对复杂的变量间关系进行建模. 偏最小二乘模型迭代选择一系列正交潜在因子, 使得潜在因子间的协方差最大化. 这样投影得到的结果在保持观测与响应具有较高相关性的同时, 还保持了较高的判别性.

假设 $\mathcal{X} \in \mathbb{R}^D$ 是 D 维的观测变量的空间, $\mathcal{Y} \in \mathbb{R}^C$ 是 C 维的观测响应变量的空间. 针对 N 个观测样本 $\mathbf{x}_i \in \mathcal{X}$, $\mathbf{y}_i \in \mathcal{Y}$; ($i = 1, \dots, N$), 偏最小二乘构造新的空间, 该空间保持了样本间的大多数变化, 同时使得观测变量与观测响应变量之间的相关性更强, 从而增强了样本的判别性亦补偿了样本的内类差异. 偏最小二乘的模型定义如下:

$$\left. \begin{aligned} \mathbf{X} &= \mathbf{V} \mathbf{P}^T + \mathbf{G} \\ \mathbf{Y} &= \mathbf{U} \mathbf{Q}^T + \mathbf{H} \end{aligned} \right\} \quad (18)$$

式中, $\mathbf{V} \in \mathbb{R}^{N \times D'}$ 和 $\mathbf{U} \in \mathbb{R}^{N \times C'}$ 是潜在因子矩阵, 因子的个数为 D' , $\mathbf{P} \in \mathbb{R}^{D \times D'}$ 和 $\mathbf{Q} \in \mathbb{R}^{C \times C'}$ 是因子装载矩阵, $\mathbf{G} \in \mathbb{R}^{N \times D}$ 和 $\mathbf{H} \in \mathbb{R}^{N \times C}$ 是残差矩阵. 因此, 偏最小二乘方法能够将原来 D 维的观测变量和 C 维的观测响应变量分别变换成 D' 维的向量. D' 亦被称作 PLS 的 component 个数.

为了获得式(18)的分解结果, 在第 i 次迭代中, 偏最小二乘方法计算两个权重向量 \mathbf{w}_i 和 \mathbf{c}_i , 使得当式(18)中的 $\mathbf{v}_i = \mathbf{X} \mathbf{w}_i$, $\mathbf{u}_i = \mathbf{Y} \mathbf{c}_i$ 时, \mathbf{X} 和 \mathbf{Y} 能保持最大的变化:

$$\left. \begin{aligned} \max_{\mathbf{w}_i} \text{Var}(\mathbf{v}_i) \\ \max_{\mathbf{c}_i} \text{Var}(\mathbf{u}_i) \end{aligned} \right\} \quad (19)$$

式中, \mathbf{v}_i 和 \mathbf{u}_i 分别表示 \mathbf{V} 和 \mathbf{U} 的第 i 个潜在投影; $\text{Var}(\cdot)$ 表示方差.

偏最小二乘在具备以上性质的同时,还要使得观测变量与观测响应变量之间保持较高的相关性:

$$\max_{w_i, c_i} [\text{Corr}(\mathbf{v}_i, \mathbf{u}_i)]^2 \quad (20)$$

$\text{Corr}(\mathbf{v}_i, \mathbf{u}_i) = \text{Cov}(\mathbf{v}_i, \mathbf{u}_i) / \sqrt{\text{Var}(\mathbf{v}_i)\text{Var}(\mathbf{u}_i)}$ 定义了 \mathbf{X} 和 \mathbf{Y} 的相关性, $\text{Cov}(\mathbf{v}_i, \mathbf{u}_i) = \mathbf{v}_i^T \mathbf{u}_i / N$ 表示样本的协方差.

偏最小二乘最终将式(19)和(20)结合起来最大化以下目标:

$$\max_{w_i, c_i} \text{Var}(\mathbf{v}_i) [\text{Corr}(\mathbf{v}_i, \mathbf{u}_i)]^2 \text{Var}(\mathbf{u}_i) \quad (21)$$

将上式等价为

$$\max_{w_i, c_i} \text{Cov}(\mathbf{v}_i, \mathbf{u}_i)^2 = \max_{w_i, c_i} \text{Cov}(\mathbf{X}\mathbf{w}_i, \mathbf{Y}\mathbf{c}_i)^2 \quad (22)$$

因此,上式通过最大化潜在空间中潜在因子的协方差,使得学习到的潜在向量具有较强的可分性,增强了该空间中样本的判别性.以上优化目标就是偏最小二乘方法中常用的非线性迭代偏最小二乘(nonlinear iterative partial least squares, NIPALS)算法^[19]的优化目标.

算法 3.1 非线性迭代偏最小二乘算法

Input: 观测变量 \mathbf{X} 和观测响应变量 \mathbf{Y} , 迭代次数 D' (PLS 的 component 个数).

Output: 权重 $\mathbf{W} \in \mathbb{R}^{D \times D'}$, $\mathbf{C} \in \mathbb{R}^{C \times D'}$; 装载矩阵 $\mathbf{P} \in \mathbb{R}^{D \times D'}$ 和 $\mathbf{Q} \in \mathbb{R}^{C \times D'}$.

Initialization: 随机初始化潜在因子向量 \mathbf{u}_i .

For $i=1, 2, \dots, D'$

依据 \mathbf{u}_i 计算观测变量的权重向量: $\mathbf{w}_i = \mathbf{X}^T \mathbf{u}_i / \mathbf{u}_i^T \mathbf{u}_i$;

将 \mathbf{w}_i 归一化: $\|\mathbf{w}_i\| = 1$;

依据 \mathbf{w}_i 计算观测变量的潜在因子向量: $\mathbf{v}_i = \mathbf{X}\mathbf{w}_i$;

依据 \mathbf{v}_i 计算观测响应变量的权重向量: $\mathbf{c}_i = \mathbf{Y}^T \mathbf{v}_i / \mathbf{v}_i^T \mathbf{v}_i$;

将 \mathbf{c}_i 归一化: $\|\mathbf{c}_i\| = 1$;

依据 \mathbf{c}_i 计算观测响应变量的潜在因子向量: $\mathbf{u}_i = \mathbf{X}\mathbf{c}_i$;

Return $\mathbf{W}, \mathbf{C}, \mathbf{P}, \mathbf{Q}$.

NIPALS 算法在第 i 次迭代中的最优解求解等价于求解以下特征值问题:

$$\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X} \mathbf{w}_i = \lambda \mathbf{w}_i \quad (23)$$

对于 \mathbf{c}_i , NIPALS 算法等价于求解以下特征值问题:

$$\mathbf{Y}^T \mathbf{X} \mathbf{X}^T \mathbf{Y} \mathbf{c}_i = \lambda \mathbf{c}_i \quad (24)$$

NIPALS 算法首先选定任意一个潜在空间,如观测响应变量所对应的潜在空间.在该空间中随机初始化潜在因子向量 \mathbf{u}_i ,再依据该初始化向量分别计算另一个潜在空间,即观测变量所对应的潜在空间的权重向量 \mathbf{w}_i 和潜在因子向量 \mathbf{v}_i ,然后利用新得到的潜在因子向量分别计算第一个潜在空间中的权重向量和 \mathbf{c}_i .为了避免过学习, NIPALS 算法迭代重复以上的过程直至算法收敛到指定的迭代次数.通常迭代次数由交叉验证获得.算法 3.1 中列出了 NIPALS 算法的详细求解过程.

3.4.2 PLS-Vector 提取

根据节 3.3 和节 3.4.1 的分析,基于偏最小二乘方法,在此提出将图像的可变性分析与图像的分类标签相结合,构造一种新的图像表达 PLS-Vector (PLSV).

首先构造给定图像的分类依赖 GMM 超向量 $\mathbf{m}(\mathbf{X})$, 假设 $\mathbf{m}(\mathbf{X})$ 和其对应的类别标签 \mathbf{y} 满足以下双线性模型:

$$\left. \begin{aligned} \mathbf{m}(\mathbf{X}) &= \mathbf{m}_0 + \mathbf{P}\mathbf{v}(\mathbf{X}) + \mathbf{g} \\ \mathbf{y} &= \mathbf{y}_0 + \mathbf{Q}\mathbf{u} + \mathbf{h} \end{aligned} \right\} \quad (25)$$

式中, \mathbf{m}_0 表示 GMM-UBM 的超向量, \mathbf{y}_0 表示观测响应变量的均值向量; $\mathbf{P} \in \mathbb{R}^{DK \times D'}$ 表示可变性矩阵, $\mathbf{Q} \in \mathbb{R}^{C \times D'}$ 表示观测响应变量对应的装载矩阵; $\mathbf{g} \in \mathbb{R}^{DK \times 1}$ 和 $\mathbf{h} \in \mathbb{R}^{C \times 1}$ 分别为观测变量与观测响应变量的残差向量; $\mathbf{v}(\mathbf{X}) \in \mathbb{R}^{D'}$ 是潜在因子.

按照式(17)的定义,将式(25)中对 $\mathbf{m}(\mathbf{X})$ 的分析定义为如下形式:

$$\left. \begin{aligned} \tilde{\mathbf{F}}(\mathbf{X}) &= \mathbf{P}\mathbf{v}(\mathbf{X}) + \mathbf{g} \\ \tilde{\mathbf{y}} &= \mathbf{Q}\mathbf{u} + \mathbf{h} \end{aligned} \right\} \quad (26)$$

式中, $\tilde{\mathbf{F}}(\mathbf{X})$ 为关于 \mathbf{X} 的白化处理后的 Baum-Welch 统计超向量.

设 $\mathcal{X} \in \mathbb{R}^{DK}$ 为节 1 中定义的图像 Baum-Welch 统计超向量空间, $\mathcal{Y} \in \mathbb{R}^C$ 为图像对应的类别空间,其中 D 和 K 分别为底层特征的维度和混合高斯模型中高斯模型的个数, C 为图像的类别个数.根据式(26)中的定义,分别将 M 幅图像的白化处理后的 Baum-Welch 统计超向量和类别标签 $\tilde{\mathbf{F}}(\mathbf{X}_i) \in \mathcal{X}$, $\mathbf{y}_i \in \mathcal{Y}$; ($i=1, \dots, M$) 按如下方式分解:

$$\left. \begin{aligned} \tilde{\mathbf{F}} &= \mathbf{V}\mathbf{P}^T + \mathbf{G} \\ \tilde{\mathbf{Y}} &= \mathbf{U}\mathbf{Q}^T + \mathbf{H} \end{aligned} \right\} \quad (27)$$

式中, $\tilde{\mathbf{F}} = [\tilde{\mathbf{F}}(X_1), \dots, \tilde{\mathbf{F}}(X_M)]^T \in \mathbb{R}^{M \times DK}$ 代表 M 幅图像的白化处理后的 Baum-Welch 统计超向量; $\tilde{\mathbf{Y}} \in \mathbb{R}^{M \times C}$ 为 M 幅图像对应的类别标签. 若 $\tilde{\mathbf{Y}}$ 的第 i 个行向量为 $\mathbf{y}_i^T = [0, 0, 1, 0, \dots, 0]^T \in \mathbb{R}^C$, 则表示第 i 幅图像属于第 3 个类别. $\mathbf{G} \in \mathbb{R}^{M \times DK}$ 和 $\mathbf{H} \in \mathbb{R}^{M \times C}$ 为残差矩阵, $\mathbf{V} \in \mathbb{R}^{M \times D'}$ 为 M 幅图像的 PLS-Vector. 对比式(18)的标准偏最小二乘模型, 式(27)具有相同的形式. 因此, 一旦按照算法 3.1 求得权重向量 $\mathbf{W} \in \mathbb{R}^{DK \times D'}$, 便可根据 $\mathbf{F} = \tilde{\mathbf{F}}\mathbf{W} \in \mathcal{F}$ 求得最终的 PLS-Vector 图像表达. 由于 $\mathcal{F} \in \mathbb{R}^{D'}$ 是低维度的潜在空间, 因此所得到的 PLS-Vector 为一种紧致图像表达. 同时, 基于提出的判别式的模型, 图像表达判别性得到了增强, 使得不同类别之间的图像区分度较强.

3.5 实验验证

为了评估所提出的判别式的图像表达模型, 下面将在著名的 PASVAL VOC 2007 (简称 VOC07)^[20] 和 Caltech 256^[21] 数据库上对文中的方法进行验证. 实验比较部分会选取当前在这两个数据库上性能最佳的方法进行比较.

3.5.1 实验设置

在验证提出的方法时, 实验首先在灰度图中以 3 个不同尺度^[7] 密集地提取 SIFT 特征^[8] 作为底层特征描述子. 再使用主成分分析将这些底层特征描述子从 128 维降至 80 维. 然后使用提取到的所有底层特征训练得到 GMM-UBM 模型, 所使用的高斯混合分量的个数为 256. 在计算 Baum-Welch 统计超向量时, 实验使用了白化处理后的一阶和二阶 Baum-Welch 统计超向量. 为了提升分类的性能, 实验采用文献[3]中提出的幂归一化的方法对白化处理后的 Baum-Welch 超向量进行归一化, 所采用的幂参数为 0.25. 所有试验均在双片 Intel Xeon X5650 2.66GHz CPU, 128GB memory 的服务器上进行.

VOC07 数据库^[20] 包含将近 1 万幅图像和 20 个物体类别. 该数据库最大的难点是图像的背景非常复杂, 同时, 图像中包含的物体尺寸、姿态、表观、形状和视角变化也非常大, 这导致图像存在很大的可变性.

Caltech 256 数据库^[21] 包含 29,780 幅图像和 256 个物体类别以及 1 个背景类别. 图像中的物体视角、姿态、形状等各异. 该数据库数据量大、类别

多, 是图像分类任务中一个极具挑战性的数据库.

为了获得最佳的分类性能, 实验中使用了空间金字塔匹配方法 (SPM)^[22] 对图像进行分块. 在 VOC07 数据库上, 使用的 SPM 分块方法是常用的 1×1 , 1×3 和 2×2 的分块方式. 在 Caltech 256 数据库上使用的分块方式是 1×1 , 2×2 和 4×4 的分块方式, 该方式被公认在 Caltech 256 数据库上具有最佳的性能, 并被多数的公开文献所采用. SPM 分块方法在构造 Baum-Welch 统计超向量时使用, 即在不同的图像分块中构造白化的 Baum-Welch 统计超向量. 具体使用方法参照了文献[3, 22].

3.5.2 实验结果与分析

在 VOC07 数据库上的实验比较了 BoW 模型中的经典方法 LLC^[5], Super Vector (SV)^[4] 和 Fisher Vector (FV)^[3]. 实验采用训练集的 2 501 幅图像和验证集的 2 510 幅图像作为训练数据, 因此训练数据共有 5 011 幅图像. 并且在测试集的 4 952 幅图像上进行测试. 该实验数据设置与文献[3-5]中的一致, 因此实验结果的比较保证了公平性. 为了得到最佳的分类性能, 在本实验设备所能承载的范围内, 试验中 LLC 采用了 25 000 个视觉关键词, SV 采用 K-means 聚类方法得到 1 024 个视觉关键词, FV 采用 GMM 聚类方法得到 256 个视觉关键词. 其中, SV 使用了一阶形式, FV 同时使用了一阶和二阶的信息. PLSV 使用了一阶和二阶的 Baum-Welch 统计信息.

表 1 列出了在 VOC07 数据库上的实验结果, 单个类别采用的是 Average Precision (AP) 的评价指标, 总的性能指标采用的是 mean Average Precision (mAP)^[23]. 实验均未采用图像翻转^[23], 分类器采用的是线性 SVM 分类器. 其中 LLC 和 SV 的实验结果取自文献[23]中的结果, 该综述文献中的结果目前为 VOC07 数据库上的主要参照结果. SV 所采用的底层特征描述子的维度是 128 维的 SIFT 特征. 本文按照文献[23]中的实验设置, 自行实现了 FV. 值得注意的是, 在文献[23]中 FV 的结果是 61.69, 该结果是在仅仅使用单一特征 SIFT 和经典 SPM^[3] 方法的条件下, 目前所有公开文献中在 VOC07 上最高的分类结果. 同时, 文章的作者提供了实验用的代码. 然而, 我们经过反复调试并未能达到文献[23]中所示的结果. 从表 1 的结果中可以看出, PLSV 获得了最佳的分类结果. 最为重要的是, 由于 PLS 的 component 的个数选择的是 30, PLSV

最终得到的特征维度仅为 30,只分别占到 LLC,SV 和 FV 的1/833,1/34 953和1/10 923.这极大地减小了分类器训练和特征存储的负担.

表 1 VOC07 数据库上的分类结果

Tab. 1 Performance on VOC07 dataset

Method	LLC	SV	FV	PLSV
codebook	25K	1 024	256	256
Feat. Dim.	25K	1 048 576	327 680	30
plane	71.05	74.32	74.27	78.85
bike	62.85	63.79	65.68	67.64
bird	47.40	47.02	51.53	53.23
boat	67.67	69.44	70.78	72.95
bottle	25.21	29.06	30.79	28.81
bus	62.70	66.46	70.10	71.98
car	77.02	77.31	77.96	80.24
cat	59.59	60.18	62.70	62.71
chair	54.24	50.19	55.52	54.22
cow	45.27	46.46	51.40	50.37
table	51.56	51.86	57.87	55.31
dog	44.24	44.07	44.95	45.64
horse	77.52	77.85	76.94	79.34
motorbike	67.05	67.12	67.71	69.50
person	83.29	83.07	83.57	84.51
plant	27.57	27.56	32.34	32.03
sheep	45.73	48.50	45.13	51.87
sofa	53.62	51.10	55.54	55.85
train	76.01	75.50	79.72	81.18
tv	52.32	52.26	57.55	55.54
mAP	57.60	58.16	60.60	61.59

鉴于当前 FV 在 VOC07 数据库上具有最佳的性能,并且在相近的视觉关键词个数下构造,实验比较了 PLSV 和 FV 在不同视觉关键词个数下的性能.如图 4 所示,当 codebook size 增大时,两种方法的性能均呈现上升的趋势.在小 codebook size 的情况下,FV 的性能要优于 PLSV.随着 codebook size 的增大,PLSV 逐渐拉近了与 FV 之间的差距,并且在 codebook size 为 128 时超过了 FV 的性能.当 codebook size 为 256 时,PLSV 的性能高出 FV 将近一个点.这表明 PLSV 在大 codebook size 的情况下要优于 FV.这一点也正好符合 PLS 的特性,即当病态问题越严重时(特征的维度远远大于样本的个数时),PLS 越能充分发挥其性能,逼近最优的结果.

为了进一步验证 PLSV 的有效性,试验在另一个主流数据库 Caltech 256 上进行了相关的比较.试验分别采用随机抽取 15,30 和 45 个样本进行训练,

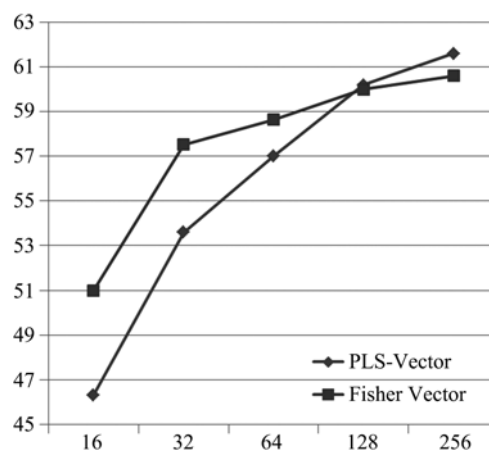


图 4 不同的 codebook size 下的分类性能

Fig. 4 Performance under different codebook size

余下的所有样本进行测试的方式.试验比较了当前在 Caltech 256 数据库上最具代表性的几个方法.为了实验结果的公正性,除了 PLSV 方法之外,其他方法的实验结果均来自原文. PLSV 一共测试了 5 次,结果为分类的精度和标准差的平均值.此处 PLS 使用的 component 个数为 20.因此,最终的 PLSV 的特征向量维度仅为 20 维.其他方法最终得到的特征向量的维度可参见各原始文献.因为多数存在的实验都采用的是单一的 SIFT 作为底层特征描述子,此处亦采用同样的底层特征进行实验.

表 2 列出了所有的实验结果. PLSV 在所有的 codebook size 下都取得了最佳的分类性能.尤其与目前综合性能最佳的 Improved Fisher Kernel(IFK)^[3]相比,PLSV 的性能取得了显著的提升.并且在训练样本数为 30 时,PLSV 的性能仍然优于当前文献中效果最好的 NN^[24]的 42.7 的实验结果,该方法使用了 5 种底层特征描述子.

表 2 Caltech 256 数据库上的分类结果

Tab. 2 Performance on Caltech 256 dataset

实验方法	训练样本数		
	15	30	45
Kernel Codebook ^[12]	—	27.2(0.4)	—
EMK (SIFT) ^[25]	23.2(0.6)	30.5(0.4)	34.4(0.4)
Standard FK (SIFT) ^[2]	25.6(0.6)	29.0(0.5)	34.9(0.2)
Sparse Coding (SIFT) ^[11]	27.7(0.5)	34.0(0.4)	37.5(0.6)
Baseline (SIFT) ^[21]	—	34.1(0.2)	—
NN (SIFT) ^[24]	—	38.0(—)	—
NN ^[24]	—	42.7(—)	—
IFK (SIFT) ^[3]	34.7(0.2)	40.8(0.1)	45.0(0.2)
PLSV (SIFT)	38.4(<0.1)	45.3(<0.1)	49.0(<0.1)

根据以上实验结果可知,PLSV 在分类性能上取得了较大的提升,其特征维度较之当前其他方法大幅度降低,减小了分类器的训练和特征存储的负担.从实验结果可以看出,首先,基于生成式模型的图像表达方法的性能优于基于重构误差最小化的方法.与 FV^[3]类似,PLSV 同样基于生成式模型进行构造,这一点要优于基于重构误差最小化的方法(比如 LLC^[5],sparse coding^[11]等方法).其次,与 FV 不同的是,PLSV 通过对图像可变性进行编码,在可变性潜在空间中对图像的充分统计量进行分析,并且融合了偏最小二乘判别性的特点,增强了图像表达的判别能力.因此,PLSV 的性能优于 FV.最后,因为 PLSV 在一个低维度的可变性潜在空间中对图像进行操作,所以 PLSV 最终获得了紧致的图像表达,这一点也是当前其他方法所不具备的,也是本文重要的贡献之一.

4 结论

基于对当前图像表达方法中存在的两个关键的局限性进行分析,本文提出了一种判别式的图像表达框架.现有方法为了获得理想的分类性能,多采用高维的特征表达,这给分类器训练和特征的存储带来了巨大的负担.同时,当前方法没有从根本上分析影响图像分类性能的因素,这导致图像表达不能抵抗图像可变性带来的负面影响.因此,本文指出存在一系列潜在的因子决定了图像的可变性,并通过可变性分析法对上述潜在因子进行分析.该方法基于偏最小二乘回归,可将每幅图像表达成一个低维的特征向量,并且增强了图像表达的判别性.在 PASCAL VOC 2007 和 Caltech 256 数据库上的实验结果表明,提出的图像表达方法在取得较好分类性能的同时,也构造了维度极低的特征向量,从而极大地降低了分类器训练和特征存储的开销.未来的工作将会在更大尺度的数据库上对 PLS-Vector 进行测试和分析.

参考文献(References)

- [1] Csurka G, Dance C R, Fan L X, et al. Visual categorization with bags of keypoints [C]// Proceedings of European Conference on Computer Vision, Prague, Czech Republic: Springer Press, 2004, 3021: 1-22.
- [2] Perronnin F, Dance C. Fisher kernels on visual vocabularies for image categorization [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minnesota, USA: IEEE Press, 2007, 1: 1-8.
- [3] Perronnin F, Sánchez J, Mensink T. Improving the Fisher kernel for large-scale image classification [C]// Proceedings of European Conference on Computer Vision, Crete, Greece: Springer, 2010, 6314: 143-156.
- [4] Zhou X, Yu K, Zhang T, et al. Image classification using super-vector coding of local image descriptors [C]// Proceedings of European Conference on Computer Vision, Crete, Greece: Springer, 2010, 6315: 141-154.
- [5] Wang J J, Yang J C, Yu K, et al. Locality-constrained linear coding for image classification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA: IEEE Press, 2010, 1: 3 360-3 367.
- [6] Huang Y Z, Huang K Q, Yu Y, et al. Salient coding for image classification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA: IEEE Press, 2011, 1: 1 753-1 760.
- [7] Zhao X, Yu Y, Huang Y Z, et al. Feature coding via vector difference for image classification [C]// Proceedings of IEEE Conference on Image Processing, Orlando, FL, USA: IEEE Press, 2012, 1: 3 121-3 124.
- [8] Lowe D G. Distinctive image visual-words from scale-invariant key-points [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [9] Mikolajczyk K, Schmid C. Scale and affine invariant interest point detectors [J]. International Journal of Computer Vision, 2004, 60(1): 63-86.
- [10] van de Sande K E A, Gevers T, Snoek C. Evaluating color descriptors for object and scene recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1 582-1 596.
- [11] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA: IEEE Press, 2009, 1: 1 794-1 801.
- [12] Gemert J C V, Geusebroek J M, Veenman C J, et al. Kernel codebooks for scene categorization [C]// Proceedings of European Conference on Computer Vision, Marseille, France: Springer, 2008, 5304: 696-709.

- [13] Perronnin F, Dance C, Csuska G, et al. Adapted vocabularies for generic visual categorization [C]// Proceedings of European Conference on Computer Vision. Graz, Austria: Springer, 2006, 3954: 464-475.
- [14] Kenny P, Ouellet P, Dehak N, et al. A study of interspeaker variability in speaker verification [J]. IEEE Transactions on Audio, Speech, Language Processing, 2008, 16(5): 980-988.
- [15] Dehak N, Kenny P J, Dehak R, et al. Front-end factor analysis for speaker verification[J]. IEEE Transactions on Audio, Speech, Language Processing, 2011, 19(4): 788-798.
- [16] West M. Bayesian factor regression model in the “large p, small n” paradigm [J]. Bayesian Statistics, 2003, 7: 723-732.
- [17] Kenny P, Boulianne G, Dumouchel P. Eigenvoice modeling with sparse training data [J] IEEE Transactions on Speech and Audio Processing, 2005, 13(3): 345-354.
- [18] Li M, Lu C, Wang A, et al. Speaker verification using Lasso based sparse total variability supervector and probabilistic linear discriminant analysis [C]// Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Hollywood, USA: IEEE Press, 2012:1-5.
- [19] Rosipal R, Krämer N. Overview and recent advances in partial least square [C]// Subspace, Latent Structure and Feature Selection Techniques. Bohinj, Slovenia: Springer, 2006, 3940: 34-51.
- [20] PASCAL2. The PASCAL Visual Object Classes [DB/OL]. <http://pascal.in.ecs.soton.ac.uk/challenges/VOC/>.
- [21] Griffin G, Holub A, Perona P. The caltech-256 [R]. Caltech Technical Report.
- [22] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2006, 2: 2 169-2 178.
- [23] Chatfield K, Lempitsky V, Vedaldi A, et al. The devil is in the details: An evaluation of recent feature encoding methods [C]// Proceedings of British Machine Vision Conference. Surrey UK: BMVA Press, 2012.
- [24] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE Press, 2008, 1: 1-8.
- [25] Bo L F, Sminchisescu C. Efficient match kernels between sets of features for visual recognition [C]// Advances in Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2009, 1: 135-143.