

# BoF 扩展模型研究

梁 晔<sup>1,2</sup> 刘宏哲<sup>1</sup> 于 剑<sup>2</sup>

(北京联合大学北京市信息服务工程重点实验室 北京 100101)<sup>1</sup>

(北京交通大学计算机与信息技术学院 北京 100044)<sup>2</sup>

**摘 要** BoF 特征是目前应用最广泛的图像表示方法。针对 BoF 特征编码简单、缺乏空间信息的缺点,对传统 BoF 流程中的特征编码和特征汇集阶段进行改进,提出了用于图像分类的新图像表示方法。首先对图像进行了基于多环划分的特征汇集的区域选择,嵌入了更多的空间信息;其次,根据密采样的特征描述子符合长尾分布的事实以及场景中特征分布比较均匀的特点,提出了适合于场景图像分类的多视觉词硬编码的编码方法。新的图像表示方法保存了 BoF 范式的优点,且特征表示更加紧凑、空间信息更加丰富。实验结果证明了所提方法的有效性。

**关键词** 特征包,特征量化,特征汇集,图像表示,图像分类

**中图法分类号** TP37 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.08.059

## Research on Extended BoF Model

LIANG Ye<sup>1,2</sup> LIU Hong-zhe<sup>1</sup> YU Jian<sup>2</sup>

(Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101, China)<sup>1</sup>

(Institute of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)<sup>2</sup>

**Abstract** BoF feature is one of the most popular image representation methods by now. Aiming at the weaknesses of hard assignment coding and discarding spatial information, the improvements of feature coding and pooling in traditional BoF paradigm were proposed. The new image representation can be used for image classification. First, multi-annulus partition method was proposed for feature pooling, which can be embedded more spatial information. Second, multi-words hard assignment coding method was proposed according to long-tail distribution of dense samples and relatively even distribution of features in scene images. The new representation not only preserves merits of BoF paradigm but also is more compact and has more spatial information. The experimental results prove the efficiency of the new method.

**Keywords** BoF, Feature quantization, Feature pooling, Image representation, Image classification

## 1 引言

随着 Internet 和多媒体技术的发展,大量的数字图像数据在丰富人们的生活、工作、教育、娱乐等方面起到越来越突出的作用。面对海量的图像数据,如何对其进行有效管理和检索变成一个紧迫的任务。图像分类技术是实现图像自动理解的一条重要途径,是一个集中了机器学习、模式识别、计算机视觉和图像处理等多个研究领域的交叉研究方向。能够自动把图像按照人们理解的方式分类到不同的语义类别,为进一步识别出图像中的其他内容提供基础,是目前研究的热点问题。

目前的图像分类方法一般而言可以分为:基于生成模型的方法和基于判别模型的方法。生成模型主要是对每个类别在特征空间中的分布分别进行建模,而判别模型则是针对不同类别在特征空间中的分布差异进行建模。在生成模型中,BoF 模型是近年来在计算机视觉领域应用最广泛的一类特征,已经应用于图像分类、对象识别、图像检索、机器人定位和

纹理识别。大量研究结果表明,BoF 特征在计算机视觉中具有很好的性能。BoF 特征是将图像表示为无序的特征集合,概念简洁、计算简单,在某些应用上取得的效果甚至可以与当前最好的方法相媲美。这种方法的缺点也很明显,即缺乏空间信息,没有考虑特征之间的关联,并且对特征分布进行了过于简单的假设。本文仔细研究了 BoF 模型,着重对 BoF 模型中的特征量编码和特征汇集阶段进行改进,并将扩展后的 BoF 模型应用于图像分类。

本文第 2 节讨论 BoF 模型的相关技术;第 3 节提出新的特征汇集和编码方法,对新算法的实现思想和实现过程进行详细阐述;第 4 节是实验结果;最后对所提方法和存在的问题进行总结。

## 2 相关工作

### 2.1 BoF 模型

BoF 特征<sup>[1]</sup>,也被称为 Bag-of-Features 或者 Bag-of-Vis-terms,其思想来源于文本信息检索和分类任务中的文档表示

到稿日期:2013-06-24 返修日期:2013-07-28 本文受国家自然科学基金项目(61271369,61372148),北京市信息服务工程重点实验室开放课题(ZK20201402),北京联合大学“新起点”计划项目(ZK201211)资助。

梁 晔(1978-),女,博士生,讲师,主要研究领域为图像处理,E-mail:liangye@buu.edu.cn(通信作者);刘宏哲(1971-),女,博士,副教授,主要研究领域为语义计算和图像处理;于 剑(1969-),男,教授,博士生导师,主要研究领域为机器学习、图像处理等。

技术,将图像表示为无序的特征集合。由于 BoF 特征是通过统计局部不变特征的全局出现情况来实现的,其特征既保留了局部特征的不变性又增加了全局特征的鲁棒性,同时与数量庞大的局部不变特征相比还能起到简化特征的作用,是对图像的压缩表示,但是丢失了特征的空间、相对位置、尺度和方向信息。构建 BoF 图像表示的过程如图 1 所示,包括特征提取、字典的生成、特征的编码和特征的汇集。

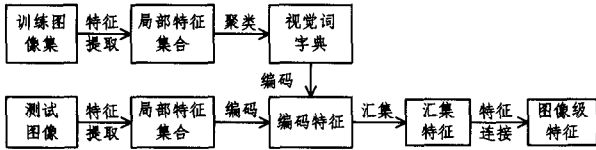


图 1 BoF 图像表示流程

BoF 表示的每个步骤的设计上都有很多选择。最近的工作<sup>[2,3]</sup>表明,对于识别任务,字典的设计相对于后面的阶段(特征编码和特征汇集)来说重要性小一些,可以通过字典对局部特征进行表示的编码阶段来获得令人满意的属性。本文着重从特征编码和汇集阶段对 BoF 模型进行扩展,并将新的编码方法和特征汇集方法应用于图像分类。

## 2.2 特征编码技术

原始的 BoF 方法<sup>[4]</sup>采用硬指派的方式对局部描述子进行编码。在生成视觉词典后,硬指派根据局部特征与视觉特征向量的近邻程度将特征描述子分配给最近的一个视觉词,被分配的视觉词对应的编码为 1,其余的视觉词的编码为 0。硬指派编码方法存在一些局限性,对字典的失真错误非常敏感;视觉单词不确定性和似然性导致的视觉词如何选择的问题。

针对硬指派编码的缺点,软指派编码中一个特征描述子用多个视觉词来描述。软指派编码的优点是概念简单、计算有效,整个计算过程不需要优化,只需要计算局部特征和每个视觉词之间的距离。稀疏编码<sup>[5,6]</sup>作为一种软指派量化的方法,已经显著提高了硬指派编码的鲁棒性。稀疏编码可以看作视觉词典的稀疏子集的线性组合,并通过  $l_1$  范式进行正则化的近似。然而,这种方法在优化的时候计算代价太大,而且还会产生相似的描述子的编码并不一致的问题<sup>[7,8]</sup>。针对稀疏编码存在的问题,K. Yu<sup>[8,9]</sup>提出了更有效、更一致的编码方法,其建立在局部性约束的基础上。作者认为描述子都位于邻近描述子的低维的流形空间内,使用欧几里得距离时将描述子分配给邻近空间内的视觉词才是有意义的,因此在编码的时候应该选择局部的基才是合理的。Lingqiao Liu<sup>[10]</sup>在传统的软分配量化编码方法的基础上加入了局部性约束,将不是  $k$  近邻的其他视觉词的距离设置为  $\infty$ ,这正是对传统的软量化的改进。显著性编码<sup>[11]</sup>是另外一种可选的办法。作者指出显著性是特征编码的基础,显著性强的视觉词应该得到更强的响应,通过特征描述子的最近邻视觉词距离和其他视觉词距离和的比值来衡量显著性,并且通过显著性进行编码,在保留简单有效的基础上显示出非常好的性能。S. Gao 于 2010 年提出了另外一种提高稀疏编码一致性的方法<sup>[7]</sup>,即在目标函数里面增加了拉普拉斯矩阵来进行字典和编码的学习。由于在数据集中存在太多的局部特征,同时构建拉普拉斯矩阵和学习稀疏编码是不可行的,计算量太大是这种方法的最大缺点。

从上面的论述看出,如何在编码的时候考虑近邻性以及

确定编码的系数仍然是研究的重点,也是本文要解决的重点问题。

## 2.3 特征汇集技术

现代的计算机视觉体系结构中通常都加入了特征的空间汇集阶段,汇集的目的就是将特征表示为一种新的更有用的特征,这种特征保留了重要信息,丢弃了不相关的细节,具有对于位置变化和光照条件变化的不变性和遮挡的鲁棒性,得到特征的压缩表示。典型的汇集操作有求和汇集(sum pooling)、求平均汇集(average pooling)和求最大汇集(max pooling)。Yang<sup>[6]</sup>在多个基准数据库上的实验结果已经显示使用特征的最大汇集操作要比平均汇集操作的结果更好。

特征汇集操作的区域选择是指在不同的图像空间内采用某种汇集操作来对特征的编码进行区域上的汇集。空间金字塔(SPM)<sup>[12]</sup>就是最经典的一种空间汇集方法。空间金字塔通过不断地将图像区域进行细分,在金字塔的子区域单元上进行特征的汇集操作而不是在整幅图像上进行特征的汇集,融入了更多的空间信息,从而大大提高了性能。空间金字塔的成功也说明了对邻域进行空间汇集操作的重要性。Hara-da T.<sup>[14]</sup>指出 SPM 在图像的每层金字塔划分时都是基于手工的且固定的划分,划分策略缺少理论依据。Yang Cao<sup>[15]</sup>从这种角度出发扩展了 BoF 模型以便包含图像中更多的空间信息。与 SPM 划分方法不同的是,首先将特征投影到某些直线或圆上来产生一系列的 BoF 特征族,使用这两种投影策略的原因是线和圆可以看成物体的基本组成元素,然后作者采用类似 boosting 的方法进行特征选择,得到最具有代表性的特征向量。这种方法可以看作是 SPM 空间思想的泛化,包含了更多的空间信息,且具有更好的平移、旋转和尺度不变性。Mateusz Malinowski<sup>[16]</sup>进一步指出了 SPM 空间划分没有考虑数据的特点和区域的硬划分的缺点,提出了联合优化分类器参数和汇集区域的方法。Yangqing Jia<sup>[17]</sup>也研究了图像汇集区域的选择对图像分类的影响,结果显示对汇集区域进行自适应的学习能够大幅提高系统的性能,即使在编码阶段使用小数量的字典。

从上面的论述看出,特征汇集中区域的选择对于分类性能的提高非常重要,如何划分新的特征汇集区域也是本文要解决的重点问题。

## 3 BoF 扩展模型设计

### 3.1 相关符号说明

图像特征编码和特征汇集的过程可以用图 2 来表示。

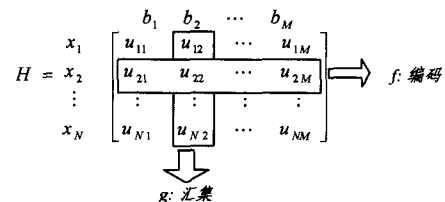


图 2 特征编码和汇集原理图

其中,  $X_{d \times N} = (x_1, x_2, \dots, x_i, \dots, x_N)$  代表局部特征描述子的集合,其中  $x_i \in R^d$ ,  $d$  表示特征的维数,  $N$  表示局部特征的个数;  $B_{d \times M} = (b_1, b_2, \dots, b_j, \dots, b_M)$  代表视觉词典,其中  $b_j \in R^d$ ,  $d$  表示特征的维数,  $M$  表示视觉词的个数;  $U = (u_1, u_2, \dots, u_N)$  代表局部特征描述子编码的集合,  $u_i$  表示特征描述子  $x_i$  的编码向量,  $u_{ij}$  表示视觉词  $b_j$  对  $x_i$  的编码。简单地

说就是需要找到两个函数  $g, f$ ,  $f$  函数负责对特征表述子进行编码,  $g$  函数负责对视觉词进行空间汇集操作, 来获得图像更紧凑、更鲁棒的表示。

### 3.2 多环划分的特征汇集区域选择方法

#### 3.2.1 算法的思想来源

SPM 是一种经典的特征汇集方法, 通过不断地将图像区域进行细分, 在金字塔的子区域单元上进行特征的汇集操作而不是在整幅图像上进行特征的汇集, 因此融入了更多的空间信息, 从而大大提高了性能。空间金字塔的成功也说明了对邻域进行空间汇集操作的重要性。SPM 已经在图像分类任务中显示出非常好的性能。由于 SPM 的空间网格划分方法的特点, 图像中的对象或区域排列大体有规律时, SPM 的分类效果较好。但是当图像中的对象排列没有规律时, SPM 的分类性能可能会有很大的下降。图 3 就是这样一个例子。在图 3 中, 菱形和黑点代表两种不同的视觉词特征。由于在两幅图像中特征的排列有明显的不同, 当通过 SPM 匹配时, 图像间的相似性会变小。可以看出, SPM 在某些情况下对图像的划分过于严格, 缺乏灵活性, 对于图像的旋转不变性和平移不变性不能很好地适应。

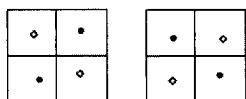


图 3 SPM 具有不好的匹配性能的例子

BOSSA<sup>[13]</sup> 是一种新的图像表示方法, 通过统计距离视觉词不同距离范围内的局部特征的个数形成对应的直方图表示, 但是这种方法实质上只考虑了特征空间的空间信息, 并没有考虑图像物理空间的信息。不过, 这种方法中按照不同距离统计时在特征空间形成的环形区域对我们有很大的启示, 能否对图像的物理空间进行环形划分是我们的出发点。

图 4 中说明对于一些图像, 如果在图像空间采用多环划分的方法, 可能会具有很高的相似性。根据图像的构图特点, 图像的主体通常分布在图像的中心, 相似的图像的周围区域也可能会有相似分布。直觉上表明多环划分的方法符合构图的原则, 能够对图像空间的特征分布进行更好的描述。

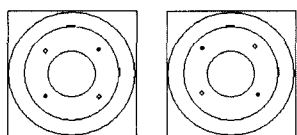


图 4 多环划分的区域选择方法能够更好地适应于旋转不变性和平移不变性

#### 3.2.2 多环划分的特征汇集区域选择方法的描述

首先, 提取每个输入图像的局部特征, 构建量化的局部特征的直方图(最上面的层  $l=1$ ); 接下来以圆环的形式逐层细化图像空间, 形成空间金字塔模型。在每一层中, 计算每个圆环区域的局部特征描述子的量化直方图, 最后, 所有的量化直方图连接起来形成图像的超级向量表示。

假设  $f_k^l \in R^M$  代表第  $l$  层的第  $k$  环内提取的特征,  $c(l)$  表示第  $l$  层中图像被划分的圆环的个数,  $L$  表示图像划分层次的总数。图像的最终特征向量为  $f, f \in R^s, s = M \sum_{l=1}^L c(l)$ , 可以

用式(1)表示。

$$f = (f_1^1, f_2^1, \dots, f_{c(2)}^2, \dots, f_1^L, \dots, f_{c(L)}^L) \quad (1)$$

图像中较细的圆环内的特征应该具有较大的权重。在本文中, 假设同一层的特征具有相同的权重。当考虑权重的时候, 会得到权重特征  $f_w, f_w \in R^s, f_w$  可以通过式(2)表示。

$$f_w = (w^1 f_1^1, w^2 f_1^1, \dots, w^2 f_{c(2)}^2, \dots, w^L f_1^L, \dots, w^L f_{c(L)}^L) \quad (2)$$

本文中, 在第 1 层中具有 1 个圆环, 在第二层中具有 2 个圆环, 第 3 层中具有 3 个圆环, 以此类推。图 5 显示了构造一个 3 层图像金字塔的方法。

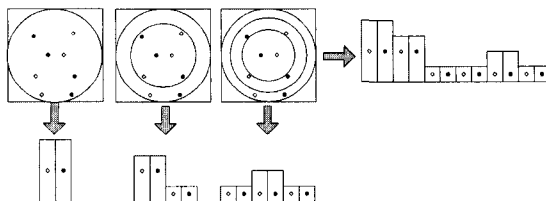


图 5 构造多环划分的 3 层金字塔的例子

在图 5 中, 图像中有两种特征, 分别表示为空心菱形和空心圆点, 逐步将图像细化为 3 层。每层的环数不一样, 分别统计每一层每一个环中归属每个视觉词的局部特征的个数, 最后将加权的空间直方图连接起来形成图像的超级向量。最后, 仍然使用空间金字塔匹配核进行匹配, 如式(3)所示, 其中  $X$  和  $Y$  代表两幅图像。

$$K(X, Y) = \sum_{l=1}^L \sum_{i=1}^{c(l)} \min(w_k^l f_i^l(X), w_k^l f_i^l(Y)) \quad (3)$$

### 3.3 多视觉词硬编码

针对硬编码存在的缺点, 许多论文提出了软编码的改进算法, 也就是一个局部特征描述子使用多个视觉词进行编码, 并且按照一定的原则给每个视觉词赋予不同的权重。但是, 在这些编码方法中编码的计算需要一定的时间花费, 计算不够简洁。研究[18]表明密采样的特征描述子的分布是符合长尾(long-tail)分布的, 也就是说特征空间实质上并没有真正的聚类中心, 每个特征都相对孤立, 因而实际上衡量每个视觉词对于编码的作用还是不能够按其欧氏距离赋予权重, 所以从这个角度上我们可以认为近邻的视觉词对于编码的作用是等同的。另外, 直觉上认为场景分布比较均匀, 没有特别突出的对象, 所以字典中近邻的视觉词对特征描述子的作用相对对象识别中的视觉词来说差异性较小。基于此, 本文提出针对场景图像的多视觉词硬编码方法。多视觉词硬编码的基本思想虽然含有软编码的思想, 但选中的视觉词被认为对于编码起相同作用, 所以选中的视觉词对应的编码值都置为 1, 没有选中的视觉词的编码设置为 0, 然后进行均一化的直方图统计。汇集操作采用平均汇集方式。多视觉词硬编码方法原理很简单, 但是应用在基于多环划分的特征汇集的区域选择之前的编码阶段却取得了比较好的效果。

编码公式可以如下:

$$u_{ij} = \begin{cases} 1, & \text{如果 } b_j \in N_k(x_i) \\ 0, & \text{否则} \end{cases}$$

$N_k(x_i)$  表示局部特征  $x_i$  的近邻视觉词。

## 4 实验

### 4.1 多环划分的特征汇集区域选择方法的实验

SPM<sup>[12]</sup>、BOSSA<sup>[13]</sup> 和 BOW 是 3 种流行的特征汇集区

域选择方法。我们在 15 场景库、Caltech-101 库、UIUC-8 库和牛津 17 花库上进行实验,比较以上 3 种方法和多环划分的特征汇集区域选择方法的分类性能。

#### 4.1.1 实验环境设置

15 场景库包括 4485 幅图像,共 15 类,每类包含 200 到 400 幅图像。图像类别既包括高山、森林等户外环境,也包括卧室和厨房等室内环境。实验中每类随机选取 100 幅图片用于训练,剩余的图片用于测试。实验在灰度范围内处理图像,采用的特征是 128 维的 SIFT<sup>[19]</sup> 特征描述子,采样间隔为 8 个像素,描述子周围区域块为  $16 \times 16$  像素。

牛津 17 花库包含 17 个类,每个类包含 80 幅图像。在实验中,每类图像中的 75% 用于训练,25% 用于测试。实验重复 10 次,每次都随机选取训练图像和测试图像,最后对 10 轮的分类率求平均和标准差。因为花的图片是彩色的,所以实验中采取 Hue-SIFT<sup>[20]</sup> 作为特征描述子。Hue-SIFT 加入了颜色信息,所以非常适合花库。实验中,提取 165 维的 Hue-SIFT 特征描述子,采样的间隔为 6 个像素。

Caltech-101 图像库包含 102 个图像类,每个类中包含 31 至 800 幅图片,共 9144 幅。每张图片的分辨率为  $300 \times 300$  像素。Caltech-101 库中图像基本都不杂乱,并且对象占据了图像的主要区域。实验中每个类随机选取 30 幅图片用于训练,剩余的图片用于测试。

UIUC-8 库包含 8 个运动类,可以用于运动分类。此图像库共包含 8 个类,每个类包含 137 至 250 幅图像,共 1579 幅。在这个数据库上,每类中随机选取 70 幅图片用来训练,60 幅图片用来测试。实验在灰度范围内处理图像,采用 128 维的 SIFT 特征描述子,采样的间隔为 8 个像素,描述子周围区域块大小为  $16 \times 16$  像素。

#### 4.1.2 实验结果

表 1 中列出了 15 场景库上的分类结果。在实验中,我们的 SPM 程序的分类结果没有文献[12]中的高,原因可能是 SIFT 描述子的提取和量化过程的差别,与此类似的情况在文献[8]中也有报道。

表 1 不同图像库的分类性能比较

|               | 多环划分的特征<br>汇集区域选择<br>方法 | SPM        | BOW        | BOSSA      |
|---------------|-------------------------|------------|------------|------------|
| 15 场景库        | 77.56±0.42              | 78.09±0.49 | 73.68±0.32 | —          |
| 牛津 17 花库      | 68.46±1.64              | 64.39±2.43 | 58.41±3.02 | 64.00±2.00 |
| Caltech-101 库 | 64.34±0.84              | 64.60±0.65 | 41.20±1.20 | —          |
| UIUC-8 库      | 81.08±1.76              | 80.42±1.44 | 78.06±1.59 | —          |

#### 4.2 多视觉词硬编码方法的实验

本文 3.3 节提出了适合于场景图像分类的多视觉词硬编码的编码方法。多视觉词硬编码方法原理很简单,但是这种编码方法和基于多环划分的特征汇集的区域选择相结合却取得了较好的分类效果。为了进一步比较分类性能,将多环划分的特征汇集区域选择方法和不同编码方法、特征汇集操作方法相结合,实验结果表明多环划分的特征区域选择、多视觉词硬编码和平均特征汇集操作相结合的分类效果最好。

实验的环境设置为:字典大小为 1000;分类器采用直方图交核的 SVM;编码方式有硬编码、多视觉词硬编码、LLC 编码<sup>[8]</sup> 和 LSC 编码<sup>[10]</sup>;特征汇集操作有求平均汇集操作和求最大值汇集操作。表 2 列出了不同情况下在 15 场景库上

的分类结果。

表 2 多环划分的特征汇集区域选择和不同编码方法、特征汇集操作相结合的分类结果

| 硬编码+<br>平均特征<br>汇集操作 | 多视觉词<br>硬编码+<br>平均特征<br>汇集操作 | LLC+<br>平均特征<br>汇集操作 | LLC+<br>最大特征<br>汇集操作 | LSC+<br>最大特征<br>汇集操作 | LSC+<br>平均特征<br>汇集操作 |
|----------------------|------------------------------|----------------------|----------------------|----------------------|----------------------|
| 77.56±<br>0.42       | 80.17±<br>0.63               | 79.08±<br>0.70       | 78.67±<br>0.52       | 77.79±<br>0.72       | 79.52±<br>0.87       |

实验结果证明了多视觉词硬编码方法在多环划分的特征汇集区域选择上是有效的。

#### 4.3 特征向量维数的讨论

本节对特征的维数进行讨论。假设字典的维数为  $M$ ,在 3 层 SPM 中,图像特征的维数为  $21M$ ,而在 3 层的多环划分图像表示中特征的维数为  $6M$ ,相同字典大小下,特征的维数不到 SPM 特征维数的  $1/3$ ;即使字典的个数不同,比如,SPM 中字典的个数为 400,基于多环划分中字典的个数为 1000,SPM 中特征的维数为  $400 \times 21 = 8400$ ,多环划分的特征维数为  $1000 \times 6 = 6000$ ,特征维数大约是 SPM 中特征维数的  $2/3$ 。随着字典维数的增加,分类精度必然也会增加,但是在多环划分的特征汇集区域选择方法下,图像特征的总维数并没有增加。实验结果证明了基于多环划分的方法在保证分类率的前提下提高了分类的速度。

**结束语** BoF 特征在计算机视觉中的应用越来越广泛<sup>[21]</sup>。本文对传统的 BoF 流程中特征编码和特征汇集阶段进行改进,提出了用于图像分类的新的图像表示方法,所提方法嵌入了图像空间的更多的空间信息。文中将提出的多环划分的特征汇集区域选择方法和其他流行的特征汇集的区域选择方法在多个图像库上进行了分类对比;此外,为了验证算法的适应性,还将多环划分的特征区域汇集方法嵌入到不同的编码方法和特征汇集操作中,实验结果表明和多视觉词硬编码的结合分类效果是最好的。这种图像表示方法概念简单且特征表示更加紧凑,保留了 BoF 范式的优点。进一步的研究方向包括:如何设置层次的数量,如何对不同层次赋予不同的权重;能否结合多特征提高分类的效率;对象图像库的特征分布和场景库具有不一样的特点,能否找到适用于对象图像库的新的编码方法并将其嵌入到多环划分的特征汇集区域选择方法中以提高分类性能,也是值得研究的方向。

#### 参考文献

- [1] Csurka G, Dance C R, Fan Li-xin, et al. Visual categorization with bags of keypoints[C]//Proceedings of European Conference Computer Vision 2004, workshop on Statistical Learning in Computer Vision. Prague, Czech Republic: Springer-Verlag LNCS, 2004: 59-74
- [2] Coates A, Ng A Y. The importance of encoding versus training with sparse coding and vector quantization[C]//ICMA. 2011
- [3] Rigamonti R, Brown M A, Lepetit V. Are sparse representations really relevant for image classification[C]//CVPR. 2011
- [4] Sivic J, Zisserman A. Video google: A Text Retrieval Approach to Object Matching in Videos[C]//Proceedings of IEEE International Conference on Computer Vision. Nice, France: IEEE Computer Society, 2003: 1470-1477
- [5] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Proceedings of Advances in Neural Information

- [6] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA; IEEE Computer Society, 2009; 1794-1801
- [7] Gao S, Tsang I, Chia L, et al. Local Features Are Not Lonely-Laplacian Sparse Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA; IEEE Computer Society, 2010; 3555-3561
- [8] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA; IEEE Computer Society, 2010; 3360-3367
- [9] Yu K, Zhang T, Gong Y. Nonlinear Learning Using Local Coordinate Coding[C]//Proceedings of Advances in Neural Information Processing System, 2009. Vancouver, British Columbia, Canada; Springer, 2009
- [10] Liu Ling-qiao, Wang Lei, Liu Xin-wang. In Defense of Soft-assignment Coding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 2486-2493
- [11] Huang Y, Huang K, Yu Y, et al. Salient Coding for Image Classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 1753-1760
- [12] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006. New York, NY, USA; IEEE Computer Society, 2006; 2169-2178
- [13] Avila S, Thome N, Cord M, et al. Bossa: Extended Bow Formalism for Image Classification[C]//Proceedings of International Conference on Image Processing, 2011. Brussels, Belgium; IEEE Computer Society, 2011; 2909-2912
- [14] Harada T, Ushiku Y, Yamashita Y, et al. Discriminative Spatial Pyramid[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2011. Colorado Springs, CO, USA; IEEE Computer Society, 2011; 1617-1624
- [15] Cao Yang, Wang Chang-hu, Li Zhi-wei, et al. Spatial-Bag-of-Features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010. San Francisco, CA, USA; IEEE Computer Society, 2010; 3352-3359
- [16] Malinowski M, Fritz M. Learnable Pooling Regions for Image Classification[C]//Proceedings of CoRR, 2013
- [17] Jia Yang-qing, Huang Chang, Darrell T. Beyond Spatial Pyramids; Receptive Field Learning for Pooled Image Features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012. Providence, RI, USA; IEEE Computer Society, 2012; 3370-3377
- [18] McCann S, Lowe D G. Local Naive Bayes Nearest Neighbor for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2012; 3650-3656
- [19] Lowe D. Distinctive Image Features from Scale-invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2); 91-110
- [20] Van De Sande, Gevers K E A T, Snoek C G M. Evaluating color descriptors for object and scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32; 1582-1596
- [21] Hui Bin, Tang Xu-sheng, Luo Hai-bo, et al. SDF Matched Filter Based on Gabor Wavelet Transform for Face Recognition[J]. Information and Control, 2008, 37(5); 633-636

(上接第 280 页)

### 参 考 文 献

- [1] Donoho D. For most large underdetermined systems of linear equations the minimal  $l^1$ -norm solution is also the sparsest solution[J]. Comm. Pure and Applied Math, 2006, 59(6); 797-829
- [2] Donoho D. For most large underdetermined systems of linear equations, the minimal  $l^1$ -norm near-solution approximates the sparsest near-solution [R]. The United States, California; Department of Statistics Stanford University, August 2004
- [3] Sivic J, zisserman A. Video google: a text retrieval approach to object matching in videos[C]//IEEE International Conference on Computer Vision (ICCV). 2003; 1470-1477
- [4] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching [C] // Proc. IEEE Conf. Computer Vision and Pattern Recognition. 2007
- [5] Lazebnik S, Schmid C, Ponce J. Beyond bags of features; spatial pyramid matching for recognizing natural scene categories[C]//Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2006
- [6] Yang Jian-chao, Yu Kai, Gong Yi-hong, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2009; 1794-1801
- [7] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//Proc. Advances in Neural Information Processing Systems 19. 2006; 801-808
- [8] Gao Sheng-hua, Tsang I W-H, Chia L-T. Laplacian sparse coding, hypergraph laplacian sparse coding, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1); 92-104
- [9] Boureau Y-L, Ponce J, LeCun Y. A Theoretical analysis of feature pooling in visual recognition[C]//Proceedings of the 27'th International Conference on machine Learning. Haifa, Israel, 2010
- [10] Boureau Y-L, Bach F, LeCun Y, et al. Learning mid-level features for recognition [C] // IEEE Conf. Computer Vision and Pattern Recognition, 2010
- [11] L Fei-fei, Fergus R, Perona P. Learning generative visual models from few training examples; an incremental bayesian approach tested on 101 object categories[C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop. 2004; 178
- [12] 王保平, 赵静, 苏建康, 等. 基于局部颜色-空间特征的图像检索方法研究[J]. 计算机应用与软件, 2012, 29(5); 28-31, 34
- [13] 兰冲. 基于统计规则的中文分词研究[D]. 西安: 西安电子科技大学, 2011