

DOI: 10.11992/tis.201602010

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.tp.20170116.1115.002.html>

一种局部聚合描述符和组显著编码相结合的编码方法

费宇杰, 吴小俊

(江南大学 物联网工程学院, 江苏 无锡 214122)

摘 要:局部聚合描述符(vector of locally aggregated descriptors, VLAD)的特征编码方法在大规模图像检索上取得了较好的效果。但是, VLAD 存在硬分配难以准确描述局部特征向量与视觉词汇隶属关系的问题, 本文将两种软分配编码与 VLAD 相结合来增强局部特征向量与视觉词汇的隶属关系。新的编码方法在 15 Scenes、Corel 10 和 UIIC Sports Event 数据库上的实验结果表明: 1) 在 VLAD 中加入局部软分配能够提高分类准确率, 而且对比 Fisher 编码在分类准确率上也有一定的优越性; 2) 除了软分配, 显著性对提高分类准确率也起到了一定的作用。

关键词:图像分类; 特征编码; 词袋; 局部聚合描述符; 软分配; 显著性

中图分类号: TP391 **文献标志码:** A **文章编号:** 1673-4785(2017)02-0172-07

中文引用格式: 费宇杰, 吴小俊. 一种局部聚合描述符和组显著编码相结合的编码方法[J]. 智能系统学报, 2017, 12(2): 172-178.

英文引用格式: FEI Yujie, WU Xiaojun. A new feature coding algorithm based on the combination of group salient coding and VLAD[J]. CAAI transactions on intelligent systems, 2017, 12(2): 172-178.

A new feature coding algorithm based on the combination of group salient coding and VLAD

FEI Yujie, WU Xiaojun

(School of IoT Engineering, Jiangnan University, Wuxi 214122, China)

Abstract: The vector of locally aggregated descriptors (VLAD) has achieved good results in addressing large-scale image retrieval problems; however, VLAD has a defect in that the relationship between local descriptors and visual words cannot be accurately described using hard assignments. In this paper, we therefore combine two kinds of soft assignment coding methods with VLAD to enhance the relationship between local feature vectors and visual words. We applied our method to 15 scenes from the Corel 10 and UIUC Sports Event datasets, with our experimental results showing that our combined partial soft assignment coding method and VLAD was able to enhance classification accuracy and achieve better classification accuracy than the well-known Fisher Coding approach. In addition to soft assignment, saliency also plays an important role in enhancing classification accuracy.

Keywords: image classification; feature coding; bag-of-features; VLAD; soft assignment; saliency

图像分类是计算机视觉和模式识别中的一个重要的研究方向, 它有广泛的应用, 例如: 视频监控^[1]、图像检索^[2]、网页内容分析^[3]。从文本分析中的 BoW (bag-of-words) 模型^[4] 发展而来的 BoF (bag-of-features) 模型^[5] 是当前最有效的图像分类

框架。

1 问题提出

如图 1 所示, BoF 模型通常包含 5 个步骤, 特征提取、字典生成、特征编码、特征池化和分类。所谓特征编码是用字典中的视觉词汇来表示图像中的局部特征向量, 局部特征在视觉词汇上的响应被称为编码系数, 将不同视觉词汇的编码系数组合在一起就是编码向量。特征编码是整个 BoF 模型的关键,

收稿日期: 2016-03-01. 网络出版日期: 2017-01-16.

基金项目: 国家自然科学基金项目 (61373055, 61672265); 江苏省教育厅科技成果转化推进项目 (JH10-28).

通信作者: 吴小俊. E-mail: xiaojun_wu_jnu@163.com.

编码的好坏会对分类效果产生巨大的影响。根据文献[6],我们将编码方法分为 4 类,如图 2 所示。编码方法中,最简单的方法是将局部特征向量指定到离它最近的视觉词汇上并设置其编码系数为非零的值,用这个视觉词汇来代表该特征向量,这种“硬指定(Hard-assignment)^[5]”的编码方法没有考虑到特征向量隶属视觉词汇的模糊性^[7],并且会产生巨大的量化误差。文献[8]提出一种“软指定(Soft-assignment)”的编码方法,通过指定特征向量到所有视觉词汇上从而减轻了“硬指定”编码带来的问题。基于重构的编码方法是选择一些视觉词汇来重构局部特征向量。例如:稀疏编码^[9](sparse coding)是使编码向量稀疏,而 LLC^[10](local-constraint linear coding)是选择局部的视觉词汇来重构特征向量。高维的编码方法,像 Fisher 核编码^[11](fisher kernel coding)和 SVC^[12](super vector coding)只需少量的视觉词汇就能获得较好的分类结果。VLAD 编码^[13]可以看作是 Fisher 核编码的简化版,下一节将对它详细介绍。最近,由于显著编码的高效性和有效性,它得到了很多关注。显著编码认为显著性是特征编码的重要特性。SaC^[14](salient coding)将显著系数作为编码系数,显著系数是通过局部特征到不同视觉词汇间的距离计算得来的。GSC^[15](group salient coding)是 SaC 的改进版,它的思想是将视觉词汇分成不同的组,不同的组得到不同的编码系数。

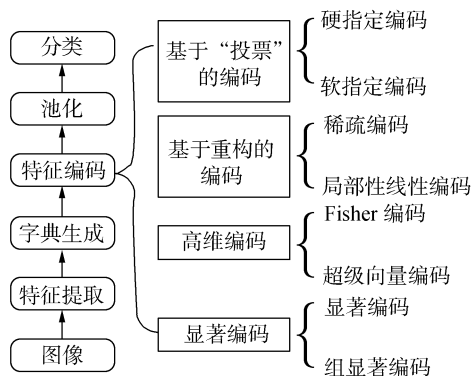


图 1 BoF 模型流程图

Fig.1 The general pipeline of the BoF framework

通过对编码方法的回顾可以发现,无论是从 Hard-assignment 到 Soft-assignment,还是从 SaC 到 GSC 都是通过软分配的思想来增加局部特征与视觉词汇之间的隶属关系信息,从而提高了分类的正确率。我们将软分配的思想加入 VLAD 中来改进原始的 VLAD。在不同的数据集上的实验结果表明:1) 软分配能够提高原始 VLAD 的分类准确率,并且对比 Fisher 核编码在分类正确率上也有一定的优越

性。但是,并不是任意的软分配都能提高分类正确率,只有局部的软分配(这里局部的意思是特征向量只有在离它最近的几个视觉词汇上有响应系数)才能提高 VLAD 的分类正确率;2) 除了软分配,显著性对提高分类正确率也起到了一定的作用,我们认为这是因为显著编码考虑到了不同视觉词汇之间的联系。

2 相关工作

在这一章节中介绍一些编码方法。 $\mathbf{b}_i (\mathbf{b}_i \in \mathbf{R}^d)$ 表示第 i 个视觉词汇; d 是局部特征向量的维数;视觉词汇的总个数为 n , 矩阵 $\mathbf{B}_{d \times n} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$ 表示视觉字典; $\mathbf{x}_i (\mathbf{x}_i \in \mathbf{R}^d)$ 表示一幅图像中第 i 个局部特征, 矩阵 $\mathbf{I}_{d \times N} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_N]$ 表示一幅图像中的 N 个局部特征; $\mathbf{u}_i (\mathbf{u}_i \in \mathbf{R}^n)$ 表示局部特征 \mathbf{x}_i 的编码向量。 u_{ij} 表示局部特征 \mathbf{x}_i 在视觉词汇 \mathbf{b}_j 上的编码系数。

1) Harding-assignment Coding。局部特征 \mathbf{x}_i 只在离它最近的视觉词汇上有编码系数。

$$u_{ij} = \begin{cases} 1, & j = \arg \min_{j=1,2,\dots,n} \|\mathbf{x}_i - \mathbf{b}_j\|_2 \\ 0, & \text{其他} \end{cases} \quad (1)$$

2) Soft-assignment Coding。 u_{ij} 可以理解为局部特征 \mathbf{x}_i 在视觉词汇 \mathbf{b}_j 上的隶属度。

$$u_{ij} = \frac{\exp(-\beta \|\mathbf{x}_i - \mathbf{b}_j\|_2)}{\sum_{k=1}^K \exp(-\beta \|\mathbf{x}_i - \mathbf{b}_k\|_2)} \quad (2)$$

β 是平滑因子控制着函数的变化率, $K=n$ 是原始的 Soft-assignment 编码,局部特征在所有的视觉词汇上都有响应系数。根据文献[16],当 K 设置为小于 n 的数时($[\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_K]$ 表示 \mathbf{x}_i 的 K 近邻视觉词汇),这种编码方法能获得更高的分类正确率。

3) Saliency Coding。显著编码(SaC)是将局部特征与离它最近的视觉词汇和其他视觉词汇的距离的比值作为该局部特征的编码结果。

$$u_{ij} = \begin{cases} \varphi(\mathbf{x}_i), & j = \arg \min_{j=1,2,\dots,n} \|\mathbf{x}_i - \mathbf{b}_j\|_2 \\ 0, & \text{其他} \end{cases} \quad (3)$$

$$\varphi(\mathbf{x}_i) = \frac{\sum_{k=2}^K (\|\mathbf{x}_i - \tilde{\mathbf{b}}_k\|_2 - \|\mathbf{x}_i - \tilde{\mathbf{b}}_1\|_2)}{\sum_{k=2}^K \|\mathbf{x}_i - \tilde{\mathbf{b}}_k\|_2} \quad (4)$$

式中: $\varphi(\mathbf{x}_i)$ 表示 \mathbf{x}_i 的显著系数, K 表示计算显著系数时所包含的视觉词汇的个数, $\tilde{\mathbf{b}}_k$ 表示离 \mathbf{x}_i 第 k 近的视觉词汇。

4) Group Saliency Coding. GSC 可以看做 SaC 的“软分配版本”,它把视觉词汇分成不同的组,局部特征在不同组上得到不同的响应系数,图 3 描绘了组显著编码的思想。

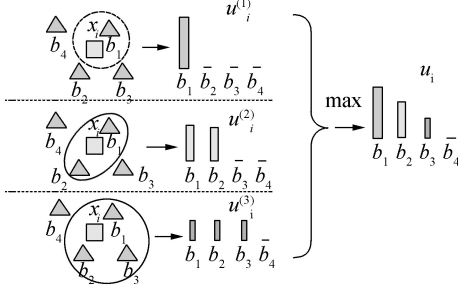


图 3 组显著编码

Fig.3 Group saliency coding

$$u_i = \max_{k=1,2,\dots,K} u_i^{(k)} \quad (5)$$

$$u_{ij}^{(k)} = \begin{cases} \varphi^{(k)}(x_i), & b_j \in g(x_i, k) \\ 0, & \text{其他} \end{cases} \quad (6)$$

$$\varphi^{(k)}(x_i) =$$

$$\sum_{t=1}^{K+1-k} (\|x_i - \tilde{b}_{k+t}\|_2 - \|x_i - \tilde{b}_k\|_2) \quad (7)$$

式中: K 表示共有 K 个组, $u_i^{(k)}$ 表示第 k 组的编码向量, $\varphi^{(k)}(x_i)$ 表示第 k 组 x_i 的显著系数, $g(x_i, k)$ 表示距离局部特征 x_i 最近的 k 个视觉词汇。

5) VLAD. VLAD 编码可以看作 Fisher 编码的简化版本, Fisher 编码的详细描述请参考文献[11], 本文不作介绍。VLAD 是将局部特征与视觉词汇的差的总和作为最后的特征向量。 $v_j (v_j \in \mathbf{R}^d)$ 表示第 j 个视觉词汇的 VLAD 编码向量。 $\mathbf{V} = [v_1^T \ v_2^T \ \dots \ v_n^T]^T (v \in \mathbf{R}^{nd})$ 是 VLAD 编码的最终形式, 表示整个图像的 VLAD 编码向量。

$$v_j = \sum_{x_i: \text{NN}(x_i) = b_j} x_i - b_j \quad (8)$$

式中: $\text{NN}(x_i) = b_j$ 表示距离局部特征 x_i 最近的视觉词汇是 b_j 。

3 改进的 VLAD 编码方法

根据上一章节的介绍, 我们发现原始的 VLAD 存在硬分配难以准确描述局部特征向量与视觉词汇隶属关系的问题。在这一章节中, 提出 3 种新的 VLAD 编码方法, 分别是 SA_VLAD 编码方法, GSC_VLAD 编码方法以及 SaC_VLAD 编码方法。

1) SA_VLAD. SA_VLAD 编码是将 Soft-assignment 编码中解决局部特征与视觉词汇隶属关系的方法加入到原始 VLAD 编码中。Soft-assignment 编码中是用高斯核函数的值来表示局部特征向量与视

觉词汇的隶属度。

$$u_{ij} = \begin{cases} \frac{\exp(-\beta \|x_i - b_j\|_2)}{\sum_{k=1}^{K_{\text{SA}}} \exp(-\beta \|x_i - b_k\|_2)}, & b_j \in g(x_i, K_{\text{SA}}) \\ 0, & \text{其他} \end{cases} \quad (9)$$

$$v_j = \sum_{i=1}^N u_{ij}(x_i - b_j) \quad (10)$$

式中: $g(x_i, K_{\text{SA}})$ 表示距离局部特征 x_i 最近的 K_{SA} 个视觉词汇。

2) GSC_VLAD. GSC_VLAD 编码是将 GSC 编码中组显著性的思想加入到原始 VLAD 中来解决局部特征向量与视觉词汇的隶属关系问题。GSC 编码中用显著性来表示局部特征向量与视觉词汇的隶属度, 同时显著性考虑到了不同视觉词汇之间的联系。

$$u_i = \max_{k=1,2,\dots,K_{\text{GSC}}} u_i^{(k)} \quad (11)$$

$$u_{ij}^{(k)} = \begin{cases} \varphi^{(k)}(x_i), & b_j \in g(x_i, k) \\ 0, & \text{其他} \end{cases} \quad (12)$$

$$\phi^{(k)}(x_i) =$$

$$\sum_{t=1}^{K_{\text{GSC}}+1-k} (\|x_i - \tilde{b}_{k+t}\|_2 - \|x_i - \tilde{b}_k\|_2) \quad (13)$$

$$v_j = \sum_{i=1}^N u_{ij}(x_i - b_j) \quad (14)$$

式中: K_{GSC} 表示有编码系数的视觉词汇的个数, 即 K_{GSC} 个视觉词汇会有响应系数。

3) SaC_VLAD. SaC 中用显著系数作为编码系数, 显著系数是指局部特征与离它最近的视觉词汇和其他视觉词汇的距离的比值, 显著系数考虑到了不同视觉词汇之间的联系。将 SaC 与原始 VLAD 相结合, 使 VLAD 在编码过程中考虑不同视觉词汇之间潜在的联系。

$$u_{ij} = \begin{cases} \varphi(x_i), & j = \arg \min_{j=1,2,\dots,n} \|x_i - b_j\|_2 \\ 0, & \text{其他} \end{cases} \quad (15)$$

$$\varphi(x_i) = \frac{\sum_{k=2}^{K_{\text{SaC}}} (\|x_i - \tilde{b}_k\|_2 - \|x_i - \tilde{b}_1\|_2)}{\sum_{k=2}^K \|x_i - \tilde{b}_k\|_2} \quad (16)$$

$$v_j = \sum_{i=1}^N u_{ij}(x_i - b_j) \quad (17)$$

式中: K_{SaC} 表示有 K_{SaC} 个视觉词汇用来计算局部特征的显著系数。

算法 1 改进的 VLAD 编码方法

输入 $I_{d \times N} = [x_1 \ x_2 \ \dots \ x_N]$ 是从图像中

得到的 N 个局部特征, $\mathbf{B}_{d \times n} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_n]$ 表示视觉字典, u_{ij} 表示局部特征 \mathbf{x}_i 在视觉词汇 \mathbf{b}_j 上的编码系数;

```
输出 VLAD 向量。
 $\mathbf{V} = [\mathbf{v}_1^T \ \mathbf{v}_2^T \ \cdots \ \mathbf{v}_n^T]^T (\mathbf{V} \in \mathbf{R}^{nd})$  % 初始化
for  $i = 1, 2, \cdots, n$ 
 $\mathbf{v}_i = \mathbf{0}_d, \mathbf{u}_i = \mathbf{0}_n$ 
end
%计算每个局部特征向量在各个视觉词汇上的
编码系数
for  $i = 1, 2, \cdots, N$ 
    根据式(9)或者式(11)或者式(15)计算相
应的  $u_{ij}$ 
end
%计算 VLAD 向量
for  $j = 1, 2, \cdots, n$ 
for  $i = 1, 2, \cdots, N$ 
 $\mathbf{v}_j = \mathbf{v}_j + u_{ij}(\mathbf{x}_i - \mathbf{b}_j)$ 
end
end
%根据文献[17],对最终的 VLAD 向量进行能
量范数和  $L_2$  范数归一化
for  $i = 1, 2, \cdots, n_d$ 
 $\mathbf{v}_i = \text{sign}(\mathbf{v}_i) \mid \mathbf{v}_i \mid^\alpha$ 
end
 $\mathbf{V} = \mathbf{V} / \|\mathbf{V}\|_2$ 
```

4 实验结果与分析

本实验为了证明以下 3 点:1) 软分配能提高 VLAD 的分类正确率,甚至对比 Fisher 编码在分类正确率上有一定的优越性;2) 只有局部软分配才能有效提高分类正确率;3) 除了软分配,显著性对提高分类正确率也起到了一定的作用。

本文在 3 个数据集(15 Scenes^[18]、Corel 10^[19]和 UIUC Sports Events^[20])上进行了实验。对于 15 Scenes 和 Corel 10 数据集,我们指定每张图片的最大单边像素为 300。UIUC Sports Events 数据集中图片的分辨率较高,因此指定每幅图片的最大单边像素为 400。我们采用 Dense SIFT 算法^[21]来提取每幅图像的 sift 特征向量,采样的步长是 6 个像素,每个采样块的大小为 16×16 。采用 K-means^[22]聚类算法生成视觉字典。分类器采用 Lib-linear SVM^[23],并指定 SVM 的惩罚系数为 1。我们将数据集分为 5 组,每组随机生成训练样本和测试样本,最后的分类正确率是 5 组实验的平均值。对于式(9)

中的 β 和 K_{SA} ,分别设置其值为 10 和 10。对于式(11)中的 K_{GSC} 设置其值为 10。关于 K_{SA} 和 K_{GSC} 对实验结果的影响,将在 4.2 节中详细讨论。

4.1 新的编码对比原始的 VLAD 和 Fisher 编码

将 4 种编码方法 SA_VLAD、GSC_VLAD、VLAD 和 Fisher 编码进行对比。

1) 15 Scenes。该数据集由 15 个场景类别构成,总共 4 485 张图片。每个类别都是相似场景图片的一个集合,大约包含 200~400 张图片,每张图片的平均尺寸为 300×250 。我们采用 Lazebnik 等^[21]的实验设置,从每类场景中随机选择 100 张图片作为训练集,其余图片作为测试集。分别在不同的视觉字典大小下进行了实验,实验结果如图 4。

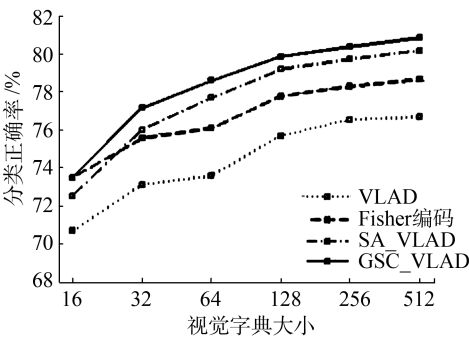


图 4 4 种不同编码方法在 15 Scenes 上的分类结果

Fig.4 Performance comparison on the 15 Scenes

从图 4 中可以看出,软分配的 VLAD 比原始的 VLAD 编码有显著的提升,当视觉字典的大小为 512 时,VLAD 的分类正确率为 76.66%,SA_VLAD 和 GSC_VLAD 的分类正确率分别为 80.18%和80.84%,有 4%~5%的提升。对比 Fisher 编码的 78.66%的正确率,也有 2%的提升。同时 GSC_VLAD 的分类正确率在不同的视觉字典大小下均高于 SA_VLAD,表 1 显示了 4 种编码各自的最佳分类正确率。

表 1 15 Scenes 数据集上的最佳分类正确率

Table 1 Best classification accuracy on 15 Scenes

编码方法	分类正确率(字典大小)
VLAD	76.66±0.19(512)
Fisher 编码	78.66±0.79(512)
SA_VLAD	80.18±0.34(512)
GSC_VLAD	80.84±0.63(512)

2) Corel 10。该数据集共有 10 个类别,每类共有 100 张图片,每张图片的平均尺寸为 384×256 。本文从每类场景中随机选择 50 张图片作为训练集,剩下的 50 张作为测试集。实验结果如图 5 所示。

实验结果基本和 15 Scenes 数据集上的结果类似,从图 4 中可以看出,SA_VLAD 和 GSC_VLAD 对比原始的 VLAD 在分类正确率上有明显的提升,并

且对比 Fisher 编码有一定的可比性。值得注意的是 GSC_VLAD 的分类正确率在不同的视觉字典大小下还是均高于 SA_VLAD。表 2 显示了 4 种编码各自的最佳分类正确率。

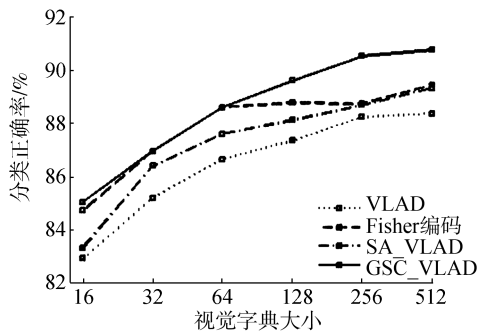


图 5 4 种不同编码方法在 Corel 10 上的分类结果
Fig.5 Performance comparison on the Corel 10

表 2 Corel10 数据集上的最佳分类正确率
Table 2 Best classification accuracy on Corel 10

编码方法	分类正确率(字典大小)
VLAD	88.36±1.05(512)
Fisher 编码	89.44±0.95(256)
SA_VLAD	90.75±0.91(512)
GSC_VLAD	89.32±0.97(512)

3) UIUC Sports Event。该数据集包含 8 个类别, 总共 1 579 张图片, 每类大约有 137 ~ 250 张图片。本文从每类中随机抽取 70 张图片作为训练集, 从余下的图片中随机抽取 60 张作为测试集。实验结果如图 6 所示。

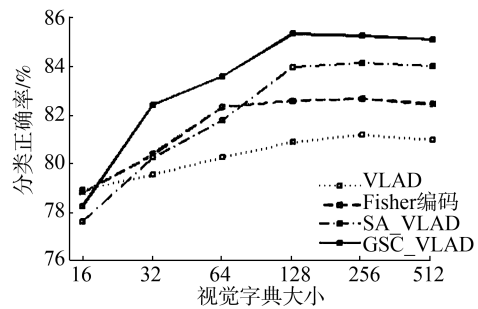


图 6 4 种不同编码方法在 UIUC 上的分类结果
Fig.6 Performance comparison on the UIUC

从图 6 中可以看出, 随着视觉字典大小的增加, SA_VLAD 和 GSC_VLAD 对比原始 VLAD 在分类正确率上有显著的增加, 对比 Fisher 编码也有一定的优越性。与上两个实验相同, GSC_VLAD 的分类正确率在不同字典大小下都超过了 SA_VLAD。表 3 显示了 4 种编码各自的最佳分类正确率。

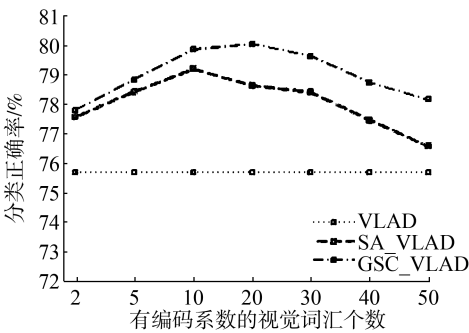
表 3 UIUC 数据集上的最佳分类正确率

Table 3 Best classification accuracy on UIUC

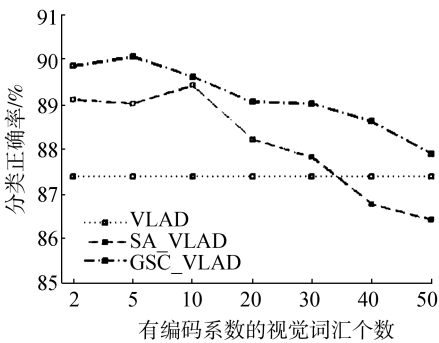
编码方法	分类正确率(字典大小)
VLAD	80.96±1.61(512)
Fisher 编码	82.67±1.06(256)
SA_VLAD	84.13±1.15(256)
GSC_VLAD	85.33±1.22(128)

4.2 算法中参数的影响

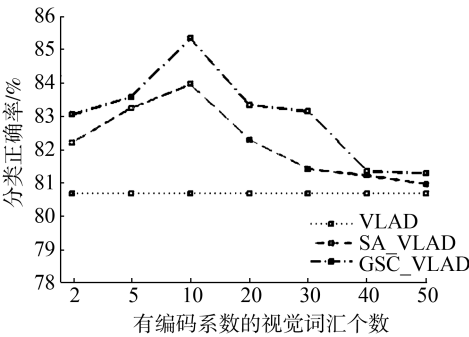
这一节中, 我们讨论式 (9) 和式 (11) 中的 K_{SA} 和 K_{GSC} 这两个参数对实验结果的影响。这两个参数表示有编码系数的视觉词汇的个数。实验中这两个参数都用 K 来表示, 当视觉字典大小为 128 时, 实验结果如图 7 所示。



(a) 15 Scenes



(b) Corel 10



(c) UIUC

图 7 参数 K 在不同数据集上的影响

Fig.7 The effect of parameter K on different dataset

图 8 中,我们加入了原始的 VLAD 以便比较。从图中可以看到随着 K 的递增,SA_VLAD 和 GSC_VLAD 的分类正确率逐渐减小,只有在 K 较小时才能达到比较高的分类正确率。这说明了并不是响应局部特征的视觉词汇越多越好,只有局部的软分配才能有效地提升 VLAD 的分类正确率。

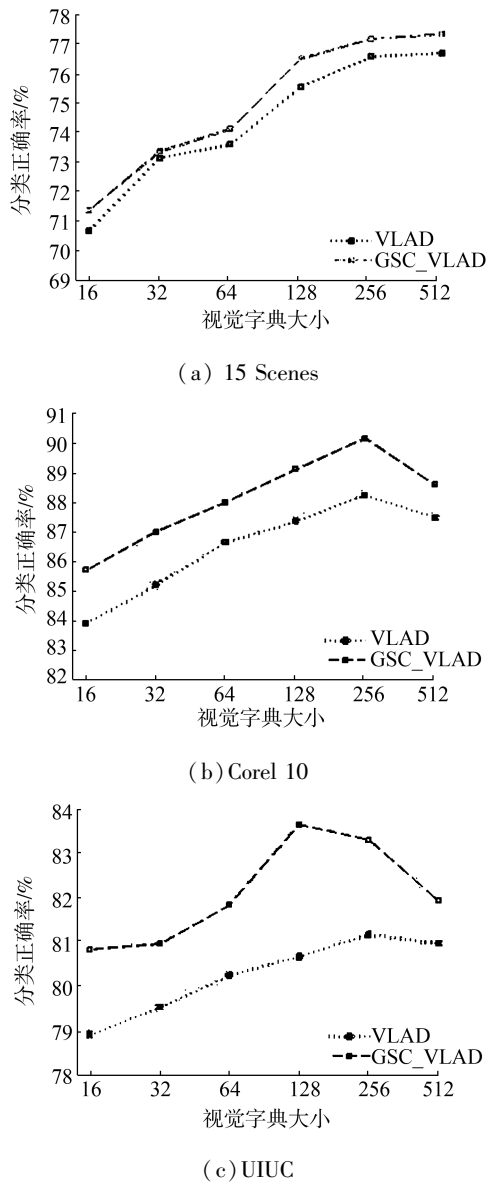


图 8 显著性对实验结果的影响
Fig.8 The effect of saliency

3.3 显著性对结果的影响

根据上面两节的介绍,我们发现 GSC_VLAD 的分类效果要略好于 SA_VLAD,我们认为这是显著性带来的影响,因为显著性考虑了不同视觉词汇间的联系。为了排除软分配对实验结果的影响,我们用新的编码算法 SaC_VLAD 与 VLAD 编码进行比较。图 7 是视觉字典大小不同时,SaC_VLAD 算法在 3 个数据集上的结果。

对比原始的 VLAD,SaC_VLAD 对分类正确率

有一定的提升。当视觉字典的大小是 128 时,在 15 Scenes 数据集上,VLAD 的分类正确率是 $75.52 \pm 0.61\%$,SaC_VLAD 的分类正确率是 $76.47 \pm 0.33\%$ 。在 Core 10 数据集上,VLAD 的分类正确率是 $87.36 \pm 1.28\%$,SaC_VLAD 的分类正确率是 $89.12 \pm 1.11\%$ 。在 UIUC 数据集上,VLAD 的分类正确率是 $80.67 \pm 1.50\%$,SaC_VLAD 的分类正确率是 $83.63 \pm 1.76\%$ 。实验结果证明,在 VLAD 中加入显著性能提高原始 VLAD 的分类正确率。

5 结束语

本文提出的 GSC_VLAD 编码方法能对原始的 VLAD 在分类正确率上带来有效的提升,同时对比 Fisher 编码也有一定的优越性。这种提升的原因是:1)在 VLAD 中加入了局部的软分配,解决了原始 VLAD 中硬分配难以准确描述局部特征向量与视觉词汇隶属关系的问题;2)显著性编码考虑了不同视觉词汇之间潜在的联系,对提升分类效果也起到了一定的作用。

在实验的过程中,我们发现新的编码方法在图像有噪声的情况下分类效果要好于原始的 VLAD,但是正确率对比没有噪声时下降得很厉害,因此如何提升新的编码方法的鲁棒性将是未来的研究重点。

参考文献:

[1] COLLINS R T, LIPTON A J, KANADE T, et al. A system for video surveillance and monitoring [R]. CMU-RITR-00-12. Pittsburgh, Penn: Carnegie Mellon University, 2000.

[2] VAILAYA A, FIGUEIREDO M A T, JAIN A K, et al. Image classification for content-based indexing [J]. IEEE transactions on image processing, 2001, 10(1): 117-130.

[3] KOSALA R, BLOCKEEL H. Web mining research: a survey [J]. ACM SIGKDD explorations newsletter, 2000, 2 (1): 1-15.

[4] JOACHIMS T. Text categorization with support vector machines: learning with many relevant features [C] // Proceedings of the 10th European Conference on Machine Learning. Berlin Heidelberg: Springer, 1998.

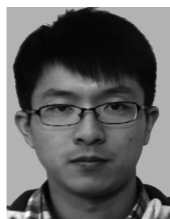
[5] DANCE C, WILLAMOWSKI J, FAN Lixin, et al. Visual categorization with bags of keypoints [C] // Proceedings of ECCV International Workshop on Statistical Learning in Computer Vision. Prague, CZ, 2004.

[6] HUANG Yongzhen, WU Zifeng, WANG Liang, et al. Feature coding in image classification: a comprehensive study [J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 36(3): 493-506.

[7] VAN GEMERT J C, VEENMAN C J, SMEULDERS A W

- M, et al. Visual word ambiguity[J]. IEEE transactions on pattern analysis and machine intelligenc, 2010, 32(7): 1271–1283.
- [8] VAN GEMERT J C, GEUSEBROEK J M, VEENMAN C J, et al. Kernel codebooks for scene categorization[C]//Proceedings of the European Conference on Computer Vision. Berlin Heidelberg: Springer, 2008: 696–709.
- [9] YANG Jianchao, YU Kai, GONG Yihong, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009: 1794–1801.
- [10] WANG Jinjun, YANG Jianchao, YU Kai, et al. Locality-constrained linear coding for image classification[C]//Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA: IEEE, 2010: 3360–3367.
- [11] PERRONNIN F, DANCE C. Fisher kernels on visual vocabularies for image categorization[C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007.
- [12] ZHOU Xi, YU Kai, ZHANG Tong, et al. Image classification using super-vector coding of local image descriptors[C]//Proceedings of the 11th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2010: 141–154.
- [13] JÉGOU H, DOUZE M, SCHMID C, et al. Aggregating local descriptors into a compact image representation[C]//Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA: IEEE, 2010: 3304–3311.
- [14] HUANG Yongzhen, HUANG Kaiqi, YU Yinan, et al. Salient coding for image classification[C]//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011.
- [15] WU Zifeng, HUANG Yongzhen, WANG Liang, et al. Group encoding of local features in image classification[C]//Proceedings of the 21st International Conference on Pattern Recognition. Tsukuba: IEEE, 2012.
- [16] LIU Lingqiao, WANG Lei, Liu Xinwang. In defense of soft-assignment coding[C]//Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona: IEEE, 2011.
- [17] PERRONNIN F, SANCHEZ J, MENSINK T. Improving the fisher kernel for large-scale image classification[C]//Proceedings of the 11th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2010: 143–156.
- [18] LI F F, PERONA P. A Bayesian hierarchical model for learning natural scene categories[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2005: 524–531.
- [19] LI Jia, WANG J Z. Automatic linguistic indexing of pictures by a statistical modeling approach[J]. IEEE transactions on pattern analysis and machine intelligence, 2003, 25(9): 1075–1088.
- [20] LI Lijia, LI Feifei. What, where and who? Classifying events by scene and object recognition[C]//Proceedings of the 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007: 1–8.
- [21] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories[C]//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA: IEEE, 2006: 2169–2178.
- [22] LLOYD S. Least squares quantization in PCM[J]. IEEE transactions on information theory, 1982, 28(2): 129–137.

作者简介:



费宇杰,男,1992年生,硕士研究生,主要研究方向为图像分类、特征编码。



吴小俊,男,1967年生,教授,主要研究方向为模式识别,计算机视觉,模糊系统,神经网络,智能系统。