

# 基于词包和特征融合的目标识别算法

周治平 李文慧 周明珠

(江南大学物联网工程学院, 无锡, 214122)

**摘要:** 针对现有词包模型对目标识别性能的不足, 对特征提取、图像表示等方面进行改进以提高目标识别的准确率。首先, 以密集提取关键点的方式取代 SIFT 关键点提取, 减少了计算时间并最大程度地描述了图像底层信息。然后采用尺度不变特征变换 (Scale-invariant feature transform, SIFT) 描述符和统一模式的局部二值模式 (Local binary pattern, LBP) 描述符描述关键点周围的形状特征和纹理特征, 引入 K-Means 聚类算法分别生成视觉词典, 然后将局部描述符进行近似局部约束线性编码, 并进行最大值特征汇聚。分别采用空间金字塔匹配生成具有空间信息的直方图, 最后将金字塔直方图相串联, 形成特征的图像级融合, 并送入 SVM 进行分类识别。在公共数据库中进行实验, 实验结果表明, 本文所提方法能取得较高的目标识别准确率。

**关键词:** 词包模型; 目标识别; 形状特征; 纹理特征

**中图分类号:** TP391      **文献标志码:** A

## Object Recognition Algorithm Based on Bag of Words and Feature Fusion

Zhou Zhiping, Li Wenhui, Zhou Mingzhu

(School of Internet of Things Engineering, Jiangnan University, Wuxi, 214122, China)

**Abstract:** For the deficiency of the existing words bag in object recognition. We improve the feature extraction and image representation etc to enhance the accuracy. Firstly, a fixed step size is used and scale-intensive is fixed to extract key points, and then the scale-invariant feature transform (SIFT) and local binry pattern(LBP) around the key points in the grids are extracted to describe the shape features and texture features. K-Means clustering algorithm is introduced to generate a visual dictionary and the local descriptors are encoded by approximated locality constrained linear coding, and max pooling and a histograms are generated using spatial pyramid matching. Both the spatial pyramid histograms are connected, therefore, the feature fusion in the image level is implemented under the words bag. Finally the fusion result is sent to the SVM for classification. Experimental result in public datasets shows that the proposed method can achieve higher recognition accuracy.

**Key words:** bag of words; object recognition; shape features; texture features

## 引 言

目标识别<sup>[1]</sup>是计算机视觉领域中的研究热点之一。当运动目标被检测出来后, 如何迅速和准确地

得到目标属性以及由此识别出目标的物体种类信息是重要的研究方向。而运动目标往往受光照、视角和尺度变化等影响,使得目标识别任务仍然面临种种挑战。传统的目标识别方法是在图像中提取特征,然后直接通过分类器对特征向量进行分类来识别目标。宋丹等<sup>[2]</sup>针对方向梯度直方图(Histogram of oriented gradient, HOG)特征不具备仿射不变性,提出仿射梯度方向直方图特征与支持向量机(Support vector machine, SVM)结合,有效抵抗了尺度变化、旋转变化和错切变化。Zhang B 等<sup>[3]</sup>提取分层梯度直方图特征与 Gabor 特征,采用最近邻分类器、多层感知器、随机树和 SVM 设计出带有拒绝选项的两级集成分类。近年来,词包模型因其简单、有效的优点而在目标识别领域得到广泛应用。基本的词包模型包括 SIFT 特征提取、生成视觉词典以及分类器实现等。为了进一步提高词包模型的识别性能,研究者们对词包的实现过程进行了不同的变种和改进<sup>[4,5]</sup>。文献[6]提取目标的轮廓分段特征,然后编码建立词包进行识别,但此方法只在二值图像上取得较好的效果,在真实图像上因其很难准确提取目标的轮廓信息而识别效果一般。文献[7]与传统的自底向上的词包构建方式不同,提取图像的中层特征,并充分利用图像的局部和全局空间信息,获得比传统的词包模型更强的判别力。文献[8]采用超像素作为分类单元,给目标的每个部件分别建立 1 个局部分类器,这在一定程度上增加了算法的复杂性。李士进等<sup>[9]</sup>提出结合相关性及冗余度分析去除视觉码本中不相关或弱相关的视觉单词,减少了计算复杂度。文献[6~9]都是采用图像的单一特征,运动目标常常受到光照、噪声、尺度以及视角变化的影响。因此单一的特征提取往往不能应对图像的多种变化,因此有必要使用多个特征进行相互补充来描述运动目标。因此许多基于特征融合的词包模型方法被提出。如文献[10]将捕捉图像形状信息的尺度不变特征变换(Scale-invariant feature transform, SIFT)特征和描述图像纹理信息的局部二值模式(Local binary pattern, LBP)特征,分别通过网格级和图像级两种方法进行特征融合。刘帅等<sup>[11]</sup>提出分别提取图像的视觉词袋特征、颜色直方图特征以及 Gabor 纹理特征,最后采用支持向量机进行分类并进行自适应综合。但文献[10,11]中算子提取兴趣点的方式都是稀疏提取的方式,会丢失掉大量的有用信息。

针对上述问题,本文从两方面进行了研究。首先在关键点提取阶段,采用固定步长、固定尺度的密集提取方式代替传统的兴趣点提取,以获取更加丰富的信息。其次,考虑到不同的图像特征描述图像基本信息的方向不一样。复杂的运动目标也使得单一特征无法充分描述其信息,提出在关键点周围提取 SIFT 和统一模式的 LBP 特征,分别采用 K-means 算法进行聚类生成视觉词典。然后将局部描述符进行近似局部约束线性编码,进行最大值特征汇聚并利用金字塔匹配生成带有空间信息的直方图并进行串联,完成形状和纹理特征的图像级融合,最后送入 SVM 进行分类识别。

## 1 特征提取和视觉词典生成

### 1.1 Dense SIFT 特征

Dense SIFT<sup>[12]</sup>特征是 SIFT 特征的一个变种,一经提出便被许多研究者使用。SIFT 算法通常分为特征点提取和特征点描述两个步骤。在特征点提取阶段,其需要构建高斯差分尺度空间等,往往耗时较长,且其提取的关键点比较稀疏,往往丢失掉过多的有用信息。Dense SIFT 采用密集提取的方法,采用固定步长,固定尺度提取大量的关键点。这样做虽然有更高的冗余度,但信息也变得丰富,再利用 SIFT 描述符描述关键点周围的形状特征能得到大量的局部描述特征,其最大程度地对图像底层进行描述,防止丢失过多有用信息。研究表明,将 Dense SIFT 与词包相结合能获得更好的性能<sup>[13]</sup>。Dense SIFT 提取的具体步骤为:从左到右、从上到下依次遍历整幅图像。每一次都提取当前位置为左上角的固定大小的网格,本文根据 Lowe 等<sup>[14]</sup>提出 SIFT 描述子的原始文献将网格大小设置为  $16 \times 16$ 。网格中心为关键点,按照一定的步长(本文设为 8)向右移动,到达图像右侧后向下移动一个步长,得到整幅图像的关键点,将提取到的关键点利用 SIFT 描述符进行描述。首先,将以关键点为中心的  $16 \times 16$  的网格划分为

$4 \times 4$  个子区域。然后在每个子区域图像小块上计算 8 个方向的梯度方向直方图,绘制每个梯度方向的累加值,即可形成一个种子点。每个种子点梯度的模值和方向分别为

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (1)$$

$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad (2)$$

式中:  $m(x, y)$  为像素点  $(x, y)$  的梯度模值;  $\theta(x, y)$  为该像素点的梯度方向,  $L(x, y)$  为该像素点的像素值。由于图像区域共存在 16 个种子点,因此每个关键点可以产生  $4 \times 4 \times 8$  共 128 维的特征向量。

## 1.2 LBP 特征

LBP 是一种用来描述图像局部纹理特征的算子,它具有旋转不变性和灰度不变性等显著优点。LBP 算子是一个固定的  $3 \times 3$  矩阵元,对应 9 个灰度值,将周边的 8 个像素灰度值与中心像素灰度值比较,大于或等于中心像素值的像素点置为 1,否则置为 0,按照逆时针或顺时针方向读取 8 个二进制值作为特征值,此后研究者对 LBP 算子进行扩充,以中心点为圆心,邻域大小为  $P$ ,从以该点半径为  $R$  的区域提取特征,记作  $LBP_{P,R}$ 。其特征为

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c) \cdot 2^i s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3)$$

式中:  $g_c$  为中心像素值;  $g_i$  为邻域像素点像素值。研究者在 LBP 的基础上提出了统一模式,实验表明最本质的纹理信息表现出强大的判别能力<sup>[15]</sup>。统一模式的 LBP 仅仅考虑  $U$  不大于 2 的情况。其中,  $U$  表示统一性度量值,为 LBP 的二进制特征值被首尾相连为环形时 0 到 1 和 1 到 0 的跳变次数。共有 58 种 LBP 统一模式,但是可以表达将近 90% 的纹理特征,而其他所有非统一模式为 1 维特征,这样就可以大大地减少特征的维数。

## 1.3 特征融合

文献[11]表明 SIFT 特征和 LBP 特征相组合对于目标识别的可行性。文献[16]将 SIFT 算法与 LBP 算法相结合,用于人脸识别,在一定程度上也提高了识别的精度。所以图像中的目标特性可以由 SIFT 和 LBP 算法共同捕获。利用密集提取的方式获得关键点之后,在关键点周围分别利用 SIFT 和统一模式的 LBP 描述符描述,并对所得结果进行归一化处理。然后对于形状特征和纹理特征分别采用 K-Means 算法进行聚类,生成视觉词典,并利用空间金字塔匹配生成带有空间信息的直方图。最后将直方图串联,实现图像级的特征融合,并送入 SVM 进行分类识别。具体识别过程如图 1 所示。

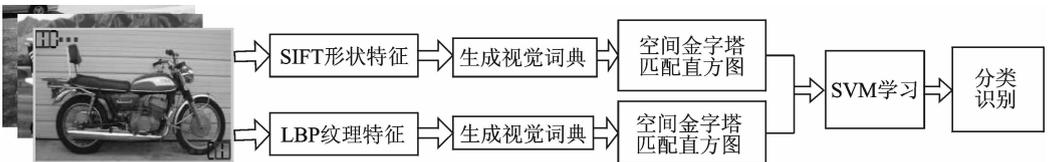


图 1 优化词包模型的目标识别过程

Fig. 1 Optimization process of bag of words

## 1.4 视觉词典生成

采用 K-Means 聚类方法对每幅图像中的特征进行聚类,将聚类的中心定义为视觉单词,然后将所有的视觉单词组合成视觉词典  $V = \{V_1, V_2, \dots, V_t\}$ 。视觉词典的大小  $t$  即为视觉单词的个数。将 SIFT 特征和 LBP 特征进行图像级融合,在融合特征中不同特征所占的比例也会直接影响最终的分类识别结果,为了获得最高的分类识别率,使用调节参数  $\phi$  来平衡各特征之间的比例,则

$$t_{LBP} = \phi t_{SIFT} \quad (4)$$

式中:  $t_s$  和  $t_{LBP}$  分别为根据 SIFT 特征和 LBP 特征学习到的视觉词典大小。

## 2 特征编码和特征汇聚

在图像中利用密集提取方式获得的大量局部特征含有很多冗余和噪声。为提高特征表示的稳定性,需要将底层特征进行编码,与无监督聚类算法得到的视觉词典相联系,得到更具区分性、更加稳定的特征表达。本文采用现阶段较为优秀的局部线性编码,并利用其近似形式,提高编码速度。具体地,近似局部线性编码的数学表达式为

$$\begin{aligned} \min_C \sum_{i=1}^N x_i - \bar{c}_i V_i^2 \\ \text{s. t. } \mathbf{I}^T \bar{\mathbf{c}}_i = 1 \quad \forall i \end{aligned} \tag{5}$$

式中:  $\mathbf{X}=[x_1, x_2, \dots, x_N]$  为局部特征向量集,  $\mathbf{C}=[c_1, c_2, \dots, c_N]$  是局部特征向量集  $\mathbf{X}$  的编码组,大小为  $t \times N$ , 每列代表每个局部特征向量在每个视觉单词上的响应值。将每个局部特征进行编码后,需要将编码后的特征集进行汇聚操作,从而形成图像内容的紧致特征表达。通过汇聚操作后的特征可以获得一定的特征不变性,同时也避免了利用整个编码特征集进行图像表达的高额计算代价。本文采用最大值汇聚方法,其数学表达式为

$$f_i = \max_{i=1,2,\dots,t} \mathbf{C}_i \tag{6}$$

取编码组  $\mathbf{C}$  每行的最大值操作,形成图像的特征表达  $\mathbf{B} = \{f_1, f_2, \dots, f_i, \dots, f_{t-1}, f_t\}$ , 并进一步对该特征向量进行  $l_2$  范数归一化操作。

## 3 空间金字塔匹配

空间金字塔划分将视觉单词的空间信息添加到词袋模型中,文献[17]中的实验证明,相比较无金字塔划分的方法,基于空间金字塔划分的方法有效提高了正确率。其基本思想为:将图像依次进行  $D+1$  层级 ( $0, 1, 2, \dots, d, \dots, D-1, D$ ) 的均匀网格划分,对于第  $d$  层,网格将图像长和宽分别划分为大小相等的  $2^d$  个单元,整幅图像即分为  $4^d$  个图像块。然后在不同层级、不同图像块中分别进行基于视觉词典的统计直方图表示。最后将所有图像块对应的统计直方图排列在一起作为整幅图像的特征。令  $\mathbf{B}_i^{d,g}$  为图像  $I$  中第  $d$  层第  $g$  个图像块所对应的直方图向量。 $\mathbf{B}_i^d$  为图像  $I$  中第  $d$  层的直方图向量,则

$$\mathbf{B}_i^d = \{\mathbf{B}_i^{d,1}, \mathbf{B}_i^{d,2}, \dots, \mathbf{B}_i^{d,g}, \dots, \mathbf{B}_i^{d,t}\} \tag{7}$$

则在第  $d$  层对统计直方图进行匹配,其相似性度量为

$$S(\mathbf{B}_i^d, \mathbf{B}_i^d) = \sum_{g=1}^{4^d} \sum_{i=1}^t \min(\mathbf{B}_i^{d,g}(i), \mathbf{B}_i^{d,g}(i)) \tag{8}$$

式中:  $\mathbf{B}_i^{d,g}(i)$  为  $\mathbf{B}_i^{d,g}$  第  $i$  维的值。方便起见,简记  $S(\mathbf{B}_i^d, \mathbf{B}_i^d)$  为  $S^d$ 。从式(5)中可以看出,第  $d+1$  层的相似性度量都被包含在第  $d$  层的相似性度量结果中,  $d=0, 1, \dots, D-1$ 。此外,注意到网格划分越少,其非匹配数目应越多,所以一般将第  $d$  层的权重设置为  $\frac{1}{2^{D-d}}$ 。其工作原理可以如图 2 所示。综合利用所有图像块,其空间金字塔匹配核可定义为

$$k = S^D + \sum_{d=0}^{D-1} \frac{1}{2^{D-d}} (S^d - S^{d+1}) = \frac{1}{2^D} S^0 + \sum_{d=1}^D \frac{1}{2^{D-d+1}} S^d \tag{9}$$

该核函数证明满足 Mercer 条件<sup>[18]</sup>,可直接用于 SVM 分类。

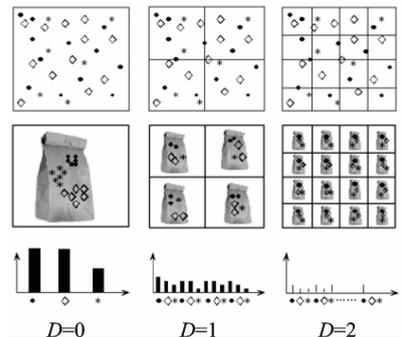


图 2 金字塔匹配原理

Fig. 2 Pyramid matching principle

## 4 实验结果及分析

### 4.1 实验数据库

本文在图像分类与识别领域中较为经典的数据库 Caltech 101 和 Caltech 256 中进行实验。图 3 为这两个数据库中的一些图像示例。Caltech 101:该数据库中一共包含 9 144 幅图像,分为 101 个物体种类和一个背景干扰类。每类中的图像数从 31~800 幅不等。与其他数据库相比,Caltech 101 数据库具有较强的颜色与形状差异,目标物体居于图像中心,且占据图像大部分面积,但缺少不同的视角变化。Caltech 256:该数据库可以看作 Caltech 101 数据库的扩展,一共包含 30 607 幅图像,分为 256 个物体种类和一个背景干扰类。每一类中至少含有 80 幅图像。此数据库不仅在物体种类和每类的图像数量上大大超过 Caltech 101 数据库,而且类内差距更加明显,类间变化更加复杂,加剧了目标分类和识别的难度。



图 3 Caltech 101 和 Caltech 256 数据库中的图像示例

Fig. 3 Image examples in Caltech 101 and Caltech 256

### 4.2 实验结果与分析

实验硬件采用 I5-3210M,主频 2.5 GHz 的处理器,8GB 内存。软件平台使用 Windows7 操作系统,开发平台为 MatlabR2012a。为了验证本文所提算法的有效性,分别在 Caltech 101 和 Caltech 256 两个数据库中进行验证。为了使实验具有说服力,分别从这两个数据库中随机选取 10 类来进行实验。这 10 类图像数据分别为:Caltech 101: Airplanes, Car\_Side, Ceiling\_fan, Dragonfly, Euphonium, Motorbikes, Nautilus, Watch, Wheelchair, Windsor\_chair。Caltech 256: Bowling-ball, Canoe, Dog, Dolphin, Flash-light, Golf-ball, School-bus, Treadmill, Yo-yo, Zebra。为了验证识别准确率与单特征 SIFT 的视觉词典大小  $t_s$  的关系,分别在上述 Caltech 101 和 Caltech 256 随机选取 10 类进行实验,利用 15 幅图像用于训练,30 幅图像用于测试,假设将图像进行 3 层级的空间金字塔匹配。得到 10 类图像的平均识别率和视觉词典大小  $t_s$  的关系,如图 4 所示。在图 4 中可以看出,上述分别在 Caltech 101 和 Caltech 256 随机选取的 10 类图像,它们的识别性能都有相似的变化趋势,即当视觉词典大小  $t$  开始增加时,平均识别率迅速增加,但当增加到一定值后继续增加,平均识别率就没有明显提高。这是因为当词典尺寸较小时,生成的视觉单词区分性不强,会导致不相似的特征映射到同一个视觉单词中。但当视觉词典大到某一数值以后继续增加,视觉单词的区分性变强了,但相似的特征可能会被映射至不同的视觉单词,使识别率上升不高甚至降低识别率。因此,视觉词典的设置并不是越大越好。从图 4,5 中还可以看出,从 Caltech 101 随机选取的 10 类图像中,其词典大小上升至 100 后,其平均识别率就上升不明显或有所下降,而在 Caltech 256 随机选取的 10 类图像中,其词典大小上升至 200 后,其平均识别率才上升不明显或有所下降,这是因为 Caltech 256 比 Caltech 101 数据库类内差距更加明显,类间变化更加复杂,因此需要比 Caltech 101 设置更大的词典大小。所以,合适的视觉词典大小需要根据实验来决定,其与所要分类的图片的种类数、图片的类内类间差距等因素息息相关。根据图 4,5 中的结果,对于上述随机选取的

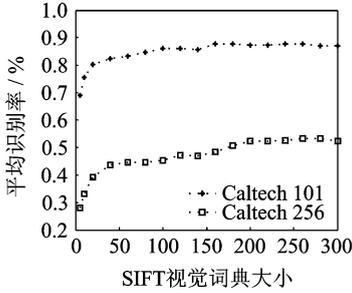


图4 目标平均识别率与单特征 SIFT 视觉词典大小的关系

Fig. 4 Relationship between object average recognition rate and size of visual dictionary of single feature SIFT

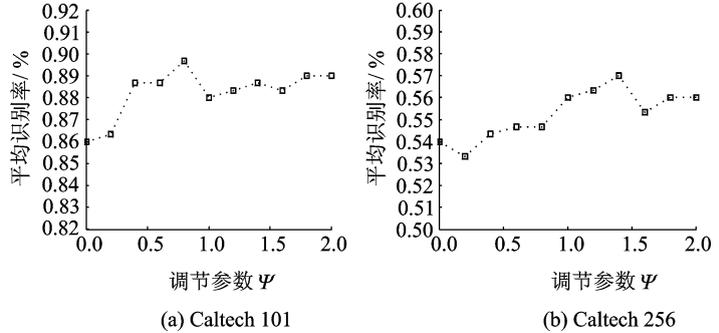


图5 目标平均识别率与调节参数  $\psi$  的关系

Fig. 5 Relationship between object average recognition rate and adjustable parameters

10 类图像中, Caltech 101 设置  $t_s$  为 100, Caltech 256 设置  $t_s$  为 200。

为了验证 SIFT 和 LBP 特征不同的组合方式对平均识别率的影响,在上述随机选取的图像中,设置不同的调节参数  $\psi$ ,得到 LBP 特征不同的视觉词典大小,将图像中利用两个特征分别得到的 3 层级划分的金子塔匹配直方图进行串联。依旧利用 15 幅图像训练,30 幅图像测试,假设将图像进行 3 层级的空间金字塔匹配。得到平均识别率与不同调节参数  $\psi$  的关系,如图 5 所示。在图 5 中可以看出,通过选取不同的调节参数  $\psi$  可以调整融合特征中 SIFT 特征和 LBP 特征各自所占的比例,并且获得不同的平均识别率。通过实验,在上述随机选取的 10 类图像中,对于 Caltech 101,当  $\psi = 0.8$  时,平均识别率最高,此时 SIFT 特征的平均识别率为 86%,LBP 特征的平均识别率为 71.33%,融合 SIFT 特征和 LBP 特征得到的平均识别率为 89.67%,比最高的单特征 SIFT 高出了 3.67%。对于 Caltech 256,当  $\psi = 1.4$  时,平均识别率最高,此时 SIFT 特征的平均识别率为 54%,LBP 特征的平均识别准确率为 38.33%,融合 SIFT 特征和 LBP 特征得到的平均识别率为 57%,比最高的单特征 SIFT 高出了 3%。为了考察金子塔匹配原理对于图像级特征融合词包模型识别性能的影响,以及划分层数如何确定。在上述随机选取的子数据库中进行实验。训练图像数设为 15,测试图像数设为 30。当  $D=0$  时代表无金子塔划分。实验结果如表 1 所示。

从表 1 中可以看出,当  $D=2$  即进行 3 层级划分时,Caltech 101 库和 Caltech 256 库的平均识别率均达到最佳,相比较无金子塔划分,平均识别率分别提高了 9.67% 和 7.67%,这充分说明了利用空间金子塔引入空间信息的必要性,同时金子塔划分的层数并非越多越好,过多的金子塔划分反而会使特征失去描述能力,降低识别率。为了验证本文所提算法的优越性,将整个数据库的实验结果与其他现有算法进行比较。根据多次实验,在整个 Caltech 101 数据库中,将  $t_s$  设置为 1 050,  $t_{LBP}$  设为 840 较为合适。将每一类的训练图像数分别设为 15 和 30,然后利用训练好的模型预测剩下的图像。在整个 Caltech 256 数据库中,将  $t_s$  设置为 2 100,  $t_{LBP}$  设为 2 940 较为合适,将每一类的训练图像数分别设为 30 和 60,然后利用训练好的模型预测剩下的图像。对比结果如表 2 所示。从表 2 可以看出,所有方法在 Caltech 256 数据库中所获得的识别性能远远小于 Caltech 101 数据库中的整体识别率,这是因为后者目标的类内变化,视角变化等更剧

表 1 基于金字塔匹配原理的平均识别率

Tab. 1 Average recognition rate based on pyramid matching

$D$	Caltech 101	Caltech 256
0	80.00	49.33
1	85.67	54.33
2	89.67	57.00
3	88.00	55.67

烈

烈,使识别任务的实现变得更加困难。对于这两个分类难度具有差异的数据库,本文所提的方法都具有一定的优势,在文献[4,5]中都只利用了形状特征,在文献[11]中虽然使用了 SIFT 和 LBP 特征,但是没有引入空间金字塔匹配原理,在一定程度上没有使得识别率达到最优。本文所提方法在 Caltech 101 数据库中当训练图像数为 30 时,整体识别率最高能达到 77.33%,对于在 Caltech 256 数据库中,当训练图像数为 60 时,整体识别率达到 50.98%。由此可见,本文方法达到了较为理想的识别效果。为了进一步验证本文所提方法的有效性,在 Graz-02 数据库,MSRcv2 物体数据集,PASCAL 2007 数据集上进行对比实验,分别得到其整体识别率。实验结果如表 3 所示。从表 3 可以看出,由于各数据集分类识别难度不同,因此获得的整体识别率也存在较大差异。但是本文算法相比较其他算法都取得了最优性能。

表 2 本文方法与现有方法整体识别率的比较

Tab. 2 Overall recognition rate comparison of proposed

比较算法	method with existing methods				%
	Caltech 101		Caltech 256		
	15	30	30	60	
本文算法	68.68	77.33	44.67	50.98	
文献[4]	65.43	73.40	41.19	47.68	
文献[5]	67.00	73.20	34.00	40.10	
文献[11]	60.50	72.86	30.50	37.60	

表 3 算法性能比较

Tab. 3 Performances comparison of algorithms %

比较算法	Graz-02	MSRcv2	PASCAL 2007	%
本文算法	60.00	92.59	49.25	
文献[4]	56.07	88.15	47.60	
文献[5]	55.33	86.72	47.48	
文献[11]	53.23	85.19	45.23	

## 5 结束语

本文分析了影响词包模型识别性能的主要因素,在特征提取与图像表示等方面对传统的词包模型进行优化。利用密集提取关键点的方式获得丰富的图像信息,并且利用形状特征和纹理特征分别描述目标的不同特性。引入 K-Means 聚类算法分别生成视觉词典,然后将局部描述符进行近似局部约束线性编码,并进行最大值特征汇聚,采用空间金字塔匹配生成带有空间信息的直方图最后送入 SVM 分类。与当前的几种流行算法比较,此方法在 Caltech,MSRcv2 等多个数据集上都表现出了良好的性能,比单一特征获得了更好的分类精度,将更好地应用在智能视频监控、车辆辅助驾驶等领域。近年来,目标识别技术越来越多地应用在实际生活中,在提高识别精度的同时减少算法的运行时间是今后研究的方向之一。

## 参考文献:

- [1] Andreopoulos A, Tsotsos J K. 50 years of object recognition: Directions forward[J]. Computer Vision and Image Understanding, 2013, 117(8): 827-891.
- [2] 宋丹, 唐林波, 赵保军. 基于仿射梯度方向直方图特征的目标识别算法[J]. 电子与信息学报, 2013, 35(6): 1428-1434.  
Song Dan, Tang Linbo, Zhao Baojun. The object recognition algorithm based on affine histogram of oriented gradient[J]. Journal of Electronics & Information Technology, 2013, 35(6): 1428-1434.
- [3] Zhang B. Reliable classification of vehicle types based on cascade classifier ensembles[J]. Intelligent Transportation Systems, IEEE Transactions on, 2013, 14(1): 322-332.
- [4] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification [C] // Proceedings of the 23rd International Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE Computer Society, 2010: 3360-3367.
- [5] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification [C] // Proceedings of the International Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society, 2009: 1794-1801.
- [6] Wang X, Feng B, Bai X, et al. Bag of contour fragments for robust shape classification[J]. Pattern Recognition, 2014, 47

(6):2116-2125.

- [7] Fernando B, Fromont E, Tuytelaars T. Mining mid-level features for image classification[J]. *International Journal of Computer Vision*, 2014, 108(3):186-203.
- [8] Lu H, Feng X, Li X, et al. Superpixel level object recognition under local learning framework [J]. *Neurocomputing*, 2013, 120:203-213.
- [9] 李士进, 仇建斌, 於慧. 基于视觉单词选择的高分辨率遥感图像飞机目标检测[J]. *数据采集与处理*, 2014, 29(1):60-65.  
Li Shijin, Qiu Jianbin, Yu Hui. Aircraft detection in high resolution remote sensing imagery based on visual words selection [J]. *Journal of Data Acquisition and Processing*, 2014, 29(1):60-65.
- [10] Yu J, Qin Z, Wan T, et al. Feature integration analysis of bag-of-features model for image retrieval[J]. *Neurocomputing*, 2013, 120(23):355-364.
- [11] 刘帅, 李士进, 冯钧. 多特征融合的遥感图像分类[J]. *数据采集与处理*, 2014, 29(1):108-115.  
Liu Shuai, Li Shijin, Feng Jun. Remote sensing image classification based on adaptive fusion of multiple features[J]. *Journal of Data Acquisition and Processing*, 2014, 29(1):108-115.
- [12] Li F F, Pietro P. A bayesian hierarchical model for learning natural scene categories[C]// *Proceedings of the International Conference on Computer Vision and Pattern Recognition*. [S. l.]: IEEE, 2005:524-531.
- [13] 赵理君, 唐婷, 霍连志, 等. 图像场景分类中视觉词包模型方法综述[J]. *中国图象图形学报*, 2014, 19(3):333-343.  
Zhao Lijun, Tang Ping, Huo Lianzhi, et al. Review of the bag-of-visual-words models in image scene classification[J]. *Journal of Image and Graphics*, 2014, 19(3):333-343.
- [14] Lowe D G. Object recognition from local scale-invariant features[C]// *Proceedings of the Seventh International Conference on Computer Vision*. Kerkyra: IEEE, 1999:1150-1157.
- [15] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7):971-987.
- [16] Yang H M, Sun J G. An improved face recognition algorithm based on SIFT and LBP[J]. *Applied Mechanics and Materials*, 2013, 427:1999-2004.
- [17] 赵春晖, 王莹. 一种基于词袋模型的图像优化分类方法[J]. *电子与信息学报*, 2012, 34(9):2064-2070.  
Zhao Chunhui, Wang Ying. An optimized method for image classification based on bag of words model[J]. *Journal of Electronics & Information Technology*, 2012, 34(9):2064-2070.
- [18] Grauman K, Darrell T. The pyramid match kernel: Discriminative classification with sets of image features [C]// *Proceedings of the International Conference on Computer Vision*. Beijing, China: IEEE, 2005:1458-1465.

#### 作者简介:



周治平(1962-),男,博士,教授,研究方向:智能检测、自动化装置和信息安全等, E-mail: zzp@jiangnan.edu.cn。



李文慧(1990-),女,硕士研究生,研究方向:视频与图像信号分析处理。



周明珠(1988-),女,硕士研究生,研究方向:检测技术与自动化装置。

