软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2012,23(7):1787−1795 [doi: 10.3724/SP.J.1001.2012.04082]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种视觉词软直方图的图像表示方法*

王彦杰[1,2], 刘峡壁[1+], 贾云得[1]

[1](北京理工大学 计算机学院 智能信息技术北京市重点实验室,北京 100081)

[2](91635 部队,北京 102249)

Visual Word Soft-Histogram for Image Representation

WANG Yan-Jie[1,2], LIU Xia-Bi[1+], JIA Yun-De[1]

[1](Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China)

[2](The 91635th Unit of PLA, Beijing 102249, China)

+ Corresponding author: E-mail: liuxiabi@bit.edu.cn, http://isc.cs.bit.edu.cn/MLMR

Wang YJ, Liu XB, Jia YD. Visual word soft-histogram for image representation. *Journal of Software*, 2012, 23(7):1787−1795 (in Chinese). http://www.jos.org.cn/1000-9825/4082.htm

Abstract: This paper proposes a visual word soft-histogram for image representation based on statistical modeling and discriminative learning of visual words. This type of learning uses Gaussian mixture models (GMM) to reflect the appearance variation of each visual word and employs the max-min posterior pseudo-probabilities discriminative learning method to estimate GMMs of visual words. The similarities between each visual word and corresponding local features are computed, summed, and normalized to construct a soft-histogram. This paper also discusses the implementation of two representation methods. The first one is called classification-based soft histogram, in which each local feature is assigned to only one visual word with maximum similarity. The second one is called completely soft histogram, in which each local feature is assigned to all the visual words. The experimental results of Caltech-4 and PASCAL VOC 2006 confirm the effectiveness of this method.

Key words: visual word; soft-histogram; image representation; Gaussian mixture model; discriminative learning

摘 要: 基于视觉词的统计建模和判别学习,提出一种视觉词软直方图的图像表示方法.假设属于同一视觉词的图像局部特征服从高斯混合分布,利用最大-最小后验伪概率判别学习方法从样本中估计该分布,计算局部特征与视觉词的相似度.累加图像中每个视觉词与对应局部特征的相似度,在全部视觉词集合上进行结果的归一化,得到图像的视觉词软直方图.讨论了两种具体实现方法:一种是基于分类的软直方图方法,该方法根据相似度最大原则建立局部特征与视觉词的对应关系;另一种是完全软直方图方法,该方法将每个局部特征匹配到所有视觉词.在数据库 Caltech-4 和 PASCAL VOC 2006 上的实验结果表明,该方法是有效的.

关键词: 视觉词;软直方图;图像表示;高斯混合模型;判别学习

中图法分类号: TP391 **文献标识码**: A

* 基金项目: 国家自然科学基金(60973059, 90920009)
收稿时间: 2011-01-13; 定稿时间: 2011-06-20

视觉词袋(bag-of-visual-words)是近年来提出的一种基于局部特征的图像表示方法,已被广泛应用于物体识别和图像检索等[1−3]领域中.该方法源于文档分析领域中的词袋(bag-of-words)表示方法[4],词袋是一种统计关键词出现频率的文档表示方法.Csurka等人[1]将词袋引入计算机视觉领域,将高维连续的图像局部特征空间进行量化,得到一组离散的具有代表性的特征向量.这些特征向量被称为视觉词(visual word),视觉词的集合被称为视觉词典(visual vocabulary).通常采用的视觉词典构建方法是利用*k*-means算法对局部特征进行聚类,每个聚类中心对应于一个视觉词.视觉词袋方法将图像中的局部特征匹配到距离最近的一个视觉词,用视觉词出现频率的直方图表示图像.根据文献[5]的定义,本文将这种表示方法称为视觉词硬直方图(visual word hard-histogram).该方法不考虑局部特征和视觉词间的相似度,所获得的图像表示不够精确[1,2].人们尝试对视觉词进行统计建模,以增强视觉词袋方法的表示能力[6,7].在视觉词统计建模的基础上,计算局部特征与视觉词间的相似度,可以提高局部特征到视觉词的匹配精度.同时,实现一个局部特征到多个视觉词的匹配,有助于减少混淆误差.我们将考虑局部特征与视觉词间相似度的视觉词图像表示称为视觉词软直方图(visual word soft-histogram).目前,主要采用高斯混合模型(Gaussian mixture model,简称GMM)对视觉词典进行统计建模,将其中的每个高斯成分作为一个视觉词.Farquhar等人[6]的工作采用了这样的思路.Perronnin[7]根据所有物体类别的局部特征集合生成一个全局GMM模型,将其中的每个高斯成分作为全局视觉词;根据每个类别的局部特征集合对全局GMM模型进行调整,将调整后的GMM模型中的每个高斯成分作为类特定的视觉词;然后再将两种视觉词合并,定义类特定直方图的图像表示.Winn等人[8]首先利用视觉词袋方法生成初始视觉词典,然后对每一类物体图像的直方图表示建立GMM模型,计算在相应视觉词典上表示图像所带来的类内紧凑性和类间可区分性.他们以最大化类内紧凑性和类间可区分性为目标,不断合并视觉词,得到规模更小且分类能力更强的视觉词典.Yang等人[9]为每一个物体类别构造一组称为视觉比特的线性分类器,计算每一个局部特征属于该类的相似度,利用图像中的所有局部特征对应的相似度计算分类损失作为分类依据.在Carneiro等人[10]的图像标注与检索方法中,利用GMM建模每一类图像的局部特征,通过计算图像中所有局部特征属于指定类别的联合概率来确定图像是否包含该类语义概念.

本文基于视觉词高斯混合建模和判别学习方法,建立视觉词软直方图的图像表示.假设属于同一视觉词的图像局部特征服从高斯混合分布,利用最大-最小后验伪概率判别学习算法[11]从样本中获得该分布,用于计算局部特征与视觉词间的相似度.累加每一视觉词与图像中对应局部特征之间的相似度,并在所有视觉词集合上进行归一化,得到图像的视觉词软直方图.根据视觉词与局部特征对应关系的不同确定方法,形成两种表示策略:基于分类的软直方图和完全软直方图.其中,基于分类的软直方图根据相似度将每一个局部特征分类到唯一的视觉词;完全软直方图则将每一个局部特征匹配到所有的视觉词.相比于已有的视觉词统计建模方法,本文方法的主要特点有:1) 对每一个视觉词进行GMM建模,而不是对整个视觉词典建模,以更准确地描述视觉词外观变化规律;2) 采用判别学习方法对视觉词统计模型中的未知参数进行学习,以提高其区分能力.

1 视觉词的统计建模与判别学习

局部特征到视觉词的匹配是视觉词袋方法的关键步骤.为了提高局部特征与视觉词的匹配精度,本文对视觉词变化规律进行高斯混合建模,利用后验伪概率分类器[11]计算局部特征到视觉词的相似度.视觉词高斯混合模型的成分个数由基于最小描述长度[12]的期望最大化算法[13]进行估计,而模型的其余未知参数则由改进的最大-最小后验伪概率判别学习(max-min posterior pseudo-probability,简称MMP)算法[11]训练得到.

1.1 高斯混合建模

一般的视觉词袋方法采用 *k*-means 算法,从图像局部特征训练集合中得到指定个数的聚类,聚类中心即为视觉词.本文在 *k*-means 聚类之后,根据每个聚类所对应的局部特征集合获得一个高斯混合模型,将该 GMM 模型作为一个视觉词,视觉词典则是一组 GMM 模型的集合.

设$\boldsymbol{x}_i$表示一个图像局部特征,$\boldsymbol{v}_j$是视觉词典中第j个视觉词,根据GMM建模结果,视觉词的类条件概率密度函数$p(\boldsymbol{x}_i|\boldsymbol{v}_j)$应服从高斯混合分布:

$$p(\boldsymbol{x}_i \mid \boldsymbol{v}_j)=\sum_{k=1}^{K} w_k N(\boldsymbol{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

其中,

$$N(\boldsymbol{x}_i \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)=(2\pi)^{-\frac{d}{2}} \mid \boldsymbol{\Sigma}_k \mid^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right) \tag{2}$$

公式(1)、公式(2)中,K为高斯成分的数量;w_k,$\boldsymbol{\mu}_k$和$\boldsymbol{\Sigma}_k$分别为第k个高斯成分的权值、均值和协方差矩阵.为了达到可以接受的计算速度,这里假设$\boldsymbol{\Sigma}_k$为对角矩阵.

图 1 所示的是视觉词 GMM 建模过程,分为局部特征提取、k-means 聚类和高斯混合建模这 3 个步骤.首先,结合局部特征检测算子和描述算子,从每个训练图像中提取一组局部特征向量;然后,在局部特征向量集合上进行 k-means 聚类;从每个聚类对应的局部特征子集中学习得到一个 GMM 模型,作为视觉词.全部视觉词构成视觉词典.

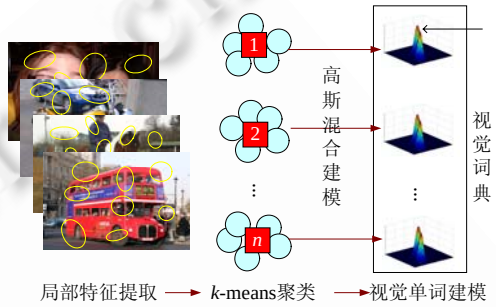


Fig.1 Statistical modeling of visual words

图 1 视觉词的统计建模过程

如前所述,一般的视觉词袋方法将聚类中心作为视觉词.与该方法相比,我们用 GMM 建模视觉词,考虑了局部特征训练数据的统计分布情况,可以用于计算局部特征与视觉词间的相似度,优于硬直方图的局部特征匹配方法.另外,相对于目前主要采用的用高斯分布建模视觉词的方法,我们用GMM建模视觉词对局部特征统计分布的描述更为精确和灵活.

1.2 后验伪概率分类器

在视觉词统计建模的基础上,利用后验伪概率分类器计算局部特征与视觉词间的相似度.定义局部特征$\boldsymbol{x}_i$属于视觉词$\boldsymbol{v}_j$的后验伪概率为

$$f(\boldsymbol{x}_i,\boldsymbol{v}_j)=1-\exp(-\lambda p^{\theta}(\boldsymbol{x}_i|\boldsymbol{v}_j)) \tag{3}$$

其中,λ和θ为正实数.该函数在以下叙述中被称为后验伪概率函数.

由公式(3)可知,后验伪概率正比于类条件概率密度,因此,后验伪概率分类器与传统的贝叶斯分类器是一致的.但进行分类决策的后验伪概率在[0,1]范围内连续取值,是局部特征与视觉词相似性的自然度量,为本文建立图像的视觉词软直方图表示奠定了基础.

1.3 视觉词模型的判别学习

在使用后验伪概率函数计算局部特征与视觉词间的相似度之前,需要确定公式(3)中的未知参数:

$$\{\lambda,\theta,w_k,\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k\},k=1,\ldots,K \tag{4}$$

利用局部特征训练集合学习上述参数的过程分为两个阶段.第 1 阶段采用基于最小描述长度的期望最大化算法,在每一视觉词对应的正样本集合上估计GMM模型中的高斯成分个数,并获得GMM模型的初始参数(包括权值w_k、均值$\boldsymbol{\mu}_k$、协方差矩阵$\boldsymbol{\Sigma}_k$);根据实验结果设置后验伪概率函数中λ和θ的初始值.第 2 阶段利用改进的MMP算法在所有样本集合(包括正样本和反样本)上修改第 1 阶段获得的初始参数.这里,某一视觉词对应的正样本是指训练集合中属于该视觉词的局部特征,而反样本则指不属于该视觉词的其他局部特征.

1.3.1 基于最小描述长度的模型选择算法

GMM模型中的成分个数是一个重要参数,对该值进行估计的问题通常被称为模型选择.本文将模型选择的最小描述长度(minimum description length,简称MDL)准则[12]与参数学习的期望最大化算法(expectation maximization,简称EM)[13]结合起来,确定GMM模型中的成分个数以及初始参数.相应的算法被称为基于最小描述长度的期望最大化算法,简称MDL-EM算法.设γ表示高斯混合模型中的所有参数个数,该值正比于模型中的高斯成分个数;$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$表示正样本数据的集合;$\boldsymbol{\Phi}$表示给定成分个数后GMM模型的参数集合,包括所有

模型成分的权值、均值和协方差矩阵; $p(\boldsymbol{x}_i \mid \boldsymbol{\Phi})_{i=1}^{n}$ 表示给定模型参数集合后,正样本数据$\boldsymbol{x}_i$对应的类条件概率密度值.令

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n \mid \boldsymbol{\Phi}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i \mid \boldsymbol{\Phi}) \tag{5}$$

则 MDL-EM 算法的学习准则是

$$\{\gamma, \boldsymbol{\Phi}\} = \arg\max_{\gamma}\left(\max_{\boldsymbol{\Phi}} \log f(\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n \mid \boldsymbol{\Phi}) - \frac{\gamma}{2}\log n\right) \tag{6}$$

其中,第 1 项表示参数学习的目标是最大化似然函数值,第 2 项表示模型选择的目标是最小化模型参数个数.因此,总体目标是在模型对数据的拟合精度和模型复杂度之间寻求平衡,以获得较好的泛化性能.MDL-EM 算法的具体流程见表 1.

Table 1 MDL-EM algorithm

表 1 MDL-EM 算法流程

Step 1. 设定成分个数的取值范围;
Step 2. 遍历其中的每一个候选成分个数:
Step 2.1. 计算 GMM 模型所包含的参数个数γ;
Step 2.2. 采用 EM 算法估计当前成分个数下的极大似然函数值;
Step 2.3. 选择使公式(6)获得最大值的候选成分个数作为当前最优模型成分个数;
Step 3. 以最优模型个数及其对应的参数集合作为结果返回.

1.3.2 改进的 MMP 参数学习

EM 算法属于生成式参数学习方法,未考虑训练集合中的反样本,在分类问题中的应用效果不如判别学习方法.本文利用 MMP 判别学习算法进一步改善 MDL-EM 算法的学习结果.MMP 算法基于后验伪概率分类器,其核心思想是,通过使正样本的后验伪概率值趋近于 1,同时使反样本的后验伪概率值趋近于 0,获得最佳分类能力.原始 MMP 算法是以类为中心的,需分别遍历训练样本集多次来学习每一类别对应的模型参数.本文学习任务所涉及的视觉词类别数量大,采用原始 MMP 算法的计算效率不够理想.因此,对原始 MMP 算法进行改进,以数据为中心来进行学习.在遍历训练样本的过程中,同时调整所有类别的模型参数,以提高计算效率.

设视觉词个数为N,$\boldsymbol{\Lambda}=\{\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2,\ldots,\boldsymbol{\Lambda}_N\}$表示所有视觉词对应的后验伪概率函数中的未知参数,$m$和$n$分别表示训练数据中某一视觉词对应的正样本和反样本个数,则改进MMP方法的目标函数为

$$F(\boldsymbol{\Lambda}) = \sum_{i=1}^{m+n}\left\{[f(\boldsymbol{x}_i;\boldsymbol{\Lambda}_\delta)-1]^2 + \sum_{j=1, j\neq\delta}^{N}[f(\boldsymbol{x}_i;\boldsymbol{\Lambda}_{v_j})]^2\right\} \tag{7}$$

在公式(7)中,$\boldsymbol{x}_i$是第i个训练样本;δ为$\boldsymbol{x}_i$所对应的视觉词序号;$f(\boldsymbol{x}_i;\boldsymbol{\Lambda}_j)$表示第$j$个视觉词对应的后验伪概率,其中,$\boldsymbol{\Lambda}_j$为第$j$个视觉词的未知参数集合.

由公式(7)可知,$F(\boldsymbol{\Lambda})$的值越小,不同类别对应的后验伪概率值的差别越大,分类效果越好;当$F(\boldsymbol{\Lambda})=0$ 时获得最佳分类效果.因此,可以通过最小化$F(\boldsymbol{\Lambda})$取得最佳参数集$\boldsymbol{\Lambda}^*$.

$$\boldsymbol{\Lambda}^* = \arg\min_{\boldsymbol{\Lambda}} F(\boldsymbol{\Lambda}) \tag{8}$$

采用梯度下降法求解公式(8)的最小值,得到最优参数集合.梯度下降法沿函数的梯度方向,迭代更新参数.设$\boldsymbol{\Lambda}^t$,α^t分别为第t次迭代时的参数集合和步长,$\nabla F(\boldsymbol{\Lambda}^t)$表示$F(\boldsymbol{\Lambda}^t)$对$\boldsymbol{\Lambda}^t$的偏导,则

$$\boldsymbol{\Lambda}^{t+1}=\boldsymbol{\Lambda}^t-\alpha^t\nabla F(\boldsymbol{\Lambda}^t) \tag{9}$$

2 视觉词软直方图

在视觉词统计建模的基础上,本文建立视觉词的软直方图来表示图像.一般的视觉词袋方法将每个局部特征匹配到唯一的视觉词,不考虑局部特征和视觉词间的相似度.然而,视觉词在图像中可能存在变化,仅考虑局部特征与视觉词之间的 0-1 关系不够鲁棒,进行匹配会导致一定的混淆误差.本文通过后验伪概率度量局部特

征和视觉词间的相似度,建立局部特征与视觉词之间的软联系.通过累计图像中视觉词对应的局部特征相似度并归一化,建立视觉词的软直方图表示,以减少匹配时的混淆误差,提高图像表示的可靠性.本文考虑了两种具体实现策略:一种策略是按照相似度大小,将局部特征匹配到一个视觉词后,建立软直方图表示.这种方法被称为基于分类的软直方图(classification based soft-histogram,简称 CBSH);另一种策略是不作分类,直接将局部特征关联到所有的视觉词.这种方法被称为完全软直方图(completely soft-histogram,简称 CSH).

设$\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_M\}$是从一个图像中提取的$M$个局部特征集合,$\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_N\}$表示包含$N$个视觉词的视觉词典,$m_i$表示被分类给视觉词$\boldsymbol{v}_i$的局部特征数量,则CBSH的计算公式为

$$\left\{\sum_{k=1}^{m_1} f(p(\boldsymbol{x}_k \mid v_1))\Big/m_1,\sum_{k=1}^{m_2} f(p(\boldsymbol{x}_k \mid v_2))\Big/m_2,\ldots,\sum_{k=1}^{m_N} f(p(\boldsymbol{x}_k \mid v_N))\Big/m_N\right\} \tag{10}$$

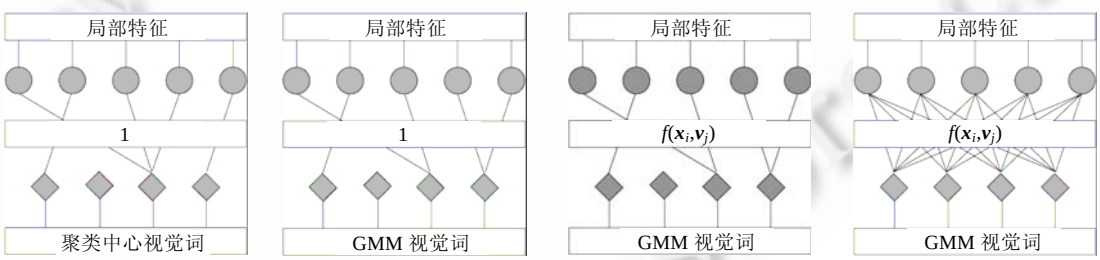
CSH 的计算公式为

$$\left\{\sum_{k=1}^{M} f(p(\boldsymbol{x}_k \mid v_1))\Big/M,\sum_{k=1}^{M} f(p(\boldsymbol{x}_k \mid v_2))\Big/M,\ldots,\sum_{k=1}^{M} f(p(\boldsymbol{x}_k \mid v_M))\Big/M\right\} \tag{11}$$

本文将所提出的视觉词软直方图表示与两种硬直方图表示进行了比较.两种硬直方图分别是基于概率的硬直方图(probability based hard-histogram,简称 PBHH)和一般视觉词袋硬直方图.其中,一般的视觉词袋计算局部特征到每个视觉词的欧几里德距离,按照距离最小化原则实现局部特征到视觉词的分类.为了便于实验比较,我们称这种方法为基于距离的硬直方图(distance based hard-histogram,简称 DBHH);PBHH 则与 CBSH 类似,利用后验伪概率分类器将局部特征分类给相应的视觉词.但与 CBSH 不同的是,在计算直方图时不考虑局部特征与视觉词的相似度,仅作 0-1 考虑.利用 PBHH 和 DBHH 得到的硬直方图均可表示为

$$\{m_1/M,m_2/M,\ldots,m_N/M\} \tag{12}$$

图 2 是两种软直方图和两种硬直方图的示意图,图中的圆形和方形分别表示局部特征和视觉词,二者之间的连线表示局部特征与视觉词间的匹配关系.



(a) DBHH (b) PBHH (c) CBSH (d) CSH

Fig.2 Comparisons of two soft-histograms and two hard-histograms

图 2 两种软直方图和两种硬直方图示意图

3 实验与讨论

本文将所提出的方法在两个数据库上进行实验验证,分别是Caltech-4[14]和PASCAL VOC 2006[15].选择Caltch-4 数据库,便于与相关算法进行比较.同时,该数据集规模较小,便于分析多种不同因素对识别结果的影响.PASCAL VOC 2006 数据库则是近些年来较为流行的物体识别数据库,包含较多的物体类别,图像也更为复杂.在获得相应的图像直方图表示之后,采用第 1.2 节所述的后验伪概率分类器和第 1.3.2 节的MMP学习算法完成物体识别任务.其中,对于物体类别的统计建模同样使用高斯混合模型.

3.1 Caltech-4数据库

3.1.1 实验设置

实验数据包括 Caltech-4 的 4 类图像(汽车尾部、人脸、自行车和飞机)和 Illinois 大学的汽车侧面图像,每

一类别的图像数量在 450~1 074 之间不等.物体大多出现在图像的中间位置,图像受光照、尺度、背景、遮挡等因素的影响较小.

采用Harris仿射不变性[16]局部特征提取算法和SIFT局部特征描述算子[17],得到 128 维局部特征矢量.与前人工作[1,14]一样,随机选择数据集中一半数量的图像作为训练集,另一半为测试集,将视觉词典中视觉词的数量设定为 1 000.

在使用 MDL-EM 算法进行模型选择时,GMM 中成分个数取值范围设定为[1,20].经计算,不同视觉词对应的成分个数在 1~5 之间发生变化,不同物体类别对应的成分个数在 3~9 之间发生变化.

3.1.2 实验结果

在本文所提出的图像表示方法中,关键因素包括视觉词的建模方法、参数学习方法以及直方图计算方法.因此,这里设计两组实验以分析不同因素对识别结果造成的影响.

第 1 组实验通过视觉词建模方法与直方图计算方法的组合,用于分析视觉词高斯混合建模方法的有效性和比较不同直方图计算方法的效果.分别基于图像的两种软直方图(CBSH,CSH)和两种硬直方图(PBHH,DBHH)进行物体识别.同时,在基于视觉词统计建模的 3 种直方图(CBSH,CSH,PBHH)表示中,分别考察了高斯模型(GM)和高斯混合模型(GMM)两种统计建模方法.相应的实验结果在表 2 中列出.由表 2 可知:1) 在CBSH,CSH 和 PBHH 这 3 种直方图表示中,高斯混合模型所对应的识别结果在绝大多数情况下优于高斯模型的识别结果,表明采用高斯混合模型相对于高斯模型能够更准确地度量局部特征和视觉词之间的相似度;2) 两种软直方图的识别结果要优于两种硬直方图.在两种软直方图之间,CSH 优于 CBSH.

Table 2 Result comparisons for visual word modeling and image representations

表 2 视觉词建模以及图像表示方法分类准确率的比较结果

Method	DBHH	PBHH		CBSH		CSH	
		GM	GMM	GM	GMM	GM	GMM
Airplane	0.961	0.968	**0.974**	0.972	**0.980**	**0.985**	0.970
Cars (rear)	0.960	0.964	**0.968**	0.945	**0.971**	0.945	**0.978**
Motor	0.889	**0.889**	**0.889**	0.893	**0.910**	0.893	**0.932**
Face	0.880	0.893	**0.907**	0.880	**0.889**	0.906	**0.933**
Cars (side)	0.953	0.942	**0.945**	0.956	**0.960**	0.956	**0.964**
Mean	0.935 9	0.938 8	**0.946 5**	0.936 4	**0.952 2**	0.943 3	**0.959 6**

第 2 组实验通过对 EM 算法和 MMP 算法学习效果的比较,验证判别学习方法的有效性.根据第 1 组实验结果,这里采用基于视觉词 GMM 模型的 CSH 方法表示图像.分别在视觉词建模和物体类别建模两个层次上考察 MMP 算法的效果,其实验结果见表 3.在两个层次上均使用 MMP 算法学习未知参数,获得了最佳的物体识别结果.

Table 3 Classification rate comparisons between MMP and EM

表 3 MMP 与 EM 对应的分类准确率比较结果

Category	物体类别 MMP 视觉词 MMP	物体类别 MMP 视觉词 EM	物体类别 EM 视觉词 MMP
Airplane	**0.980**	0.977	0.970
Cars (rear)	**0.995**	0.980	0.978
Motorbike	**0.960**	0.945	0.932
Face	**0.947**	0.933	0.933
Cars (side)	**1.000**	0.970	0.964
Mean	**0.977 3**	0.961 0	0.959 6

表 2、表 3 表明:选择视觉词GMM模型、CSH直方图计算方法以及MMP参数学习方法时,本文所提出的算法获得了最佳识别结果.我们将该结果与近期发表的相关工作进行了比较,比较结果见表 4.Csurka等人[1]和Sivic等人[18]的算法采用与本文算法相同的局部特征,即Harris仿射不变性检测算子和SIFT描述算子;Opelt等人[19]实验了两种局部特征,其中一种是Harris仿射不变性检测算子和矩不变性描述算子,该局部特征类似于本文所采用的局部特征.与上述两种采用相同或相似局部特征的算法相比,本文算法在所有物体类别的识别实验中

取得了更高的识别准确率.Fergus等人[14]的局部特征提取采用Kadir-Brady检测算子和基于像素的描述算子.本文算法在Airplane,Car(rear)和Motor这 3 个类别上的识别结果优于Fergus等人的算法.Kapoor等人[20]对于Motor和Face类别的实验结果较为优越,但他们应用了基于多分辨率的局部特征,比本文采用的局部特征要复杂.

Table 4 Comparisons with related work

表 4 与相关算法的比较结果

Category	Ours	Csurka[1]	Opelt[19]	Fergus[14]	Sivic[18]	Kapoor[20]
Airplane	0.980	0.963	0.889	0.902	0.953	0.980
Car (rear)	0.995	0.977	0.911	0.900	0.981	0.991
Motor	0.960	0.927	0.922	0.925	0.836	0.97
Face	0.947	0.94	0.935	0.964	0.940	0.995
Car (side)	1.000	0.996	0.830	—	—	—
Mean	0.977	0.961	0.897	—	—	—

3.2 PASCAL VOC 2006数据库

3.2.1 实验设置

PASCAL VOC 2006 数据库为 2006 年的PASCAL VOC Challenge竞赛所采用的图像数据集[15],目前已在物体识别领域被广泛使用,以验证物体识别算法.PASCAL VOC 2006 数据库包含 10 个物体类别:自行车(bicycle)、公共汽车(bus)、小汽车(car)、猫(cat)、牛(cow)、狗(dog)、马(horse)、摩托车(motorbike)、人(person)和羊(sheep).图像总数为 5 304,其中,2 618 幅图像用于训练,其余 2 686 幅用于测试.所包含的图像均来自自然场景图像,由于受到光照条件、拍摄角度、遮挡、背景混淆等因素的影响,同类别图像具有较大差异.

本文采用 5 个不同尺度的规则网格特征和 SIFT 描述算子,得到 128 维的局部特征矢量.为了提高算法的计算效率,采用主成分分析算法(PCA)将特征降至 50 维.根据实验结果,设定视觉词典中视觉词的数量为 3 000.

在使用 MDL-EM 算法进行模型选择时,GMM 中成分个数取值范围设定为[1,20].经计算,不同视觉词对应的成分个数在 2~7 之间变化,不同物体类别对应的成分个数在 6~15 之间发生变化.

3.2.2 实验结果

在PASCAL VOC 2006 竞赛中,识别算法性能通过接受器操作特性(receiver operating characteristic,简称ROC)曲线进行评价,具体量化评测值是ROC曲线下的区域面积(area under curve,简称AUC)[15],该值反映了分类器的平均分类性能.平均AUC值越大,分类器分类性能就越好.本文同样通过平均AUC值来比较基于不同图像表示方法的物体识别算法的性能.

与 Caltech-4 上的物体类别识别实验相同,我们在相同的条件下比较两种软直方图与两种硬直方图对应的物体识别效果,其结果见表 5.与 Caltech-4 实验相同,在 PASCAL VOC 2006 上的实验数据反映了如下实验结果:1) 两种软直方图对应的识别结果优于两种硬直方图.CSH 和 CBSH 分别将视觉词袋对应的平均 AUC 从 0.883 提高到 0.913 和 0.903,分别提高了 3.4%和 2.3%;2) 在两种软直方图之间,CSH 优于 CBSH.

Table 5 AUC comparisons between soft-histograms and hard-histograms

表 5 两种软直方图与两种硬直方图的 AUC 比较结果

AUC	Bike	Bus	Car	Cat	Cow	Dog	Horse	Motor	Person	Sheep	Average
CSH	0.932	0.977	0.959	0.924	0.928	0.837	0.891	0.921	0.792	0.925	0.913
CBSH	0.911	0.970	0.945	0.920	0.916	0.821	0.880	0.917	0.790	0.917	0.903
PBHH	0.911	0.965	0.940	0.890	0.911	0.783	0.880	0.905	0.764	0.912	0.890
DBHH	0.909	0.947	0.933	0.883	0.895	0.783	0.880	0.898	0.755	0.902	0.883

在 2006 年的物体类别识别竞赛中,竞赛者使用 20 种方法参加了全部 10 个类的挑战[15].本文利用CSH图像表示方法,取得的 10 类物体平均AUC是 0.913,该值优于参加PASCAL VOC 2006 竞赛的 16 种方法所给出的结果,但弱于最好的 4 个结果,详细比较如图 3 所示.本文工作的重点在于视觉词统计建模及其图像表示方法,没有过多关注低层局部特征提取与融合问题,仅采用了一种相对比较简单的局部特征检测和描述算子.而参与竞赛

的很多方法融合了多种局部特征检测和描述算子,包括图 3 中所示优于本文方法的 4 种算法.

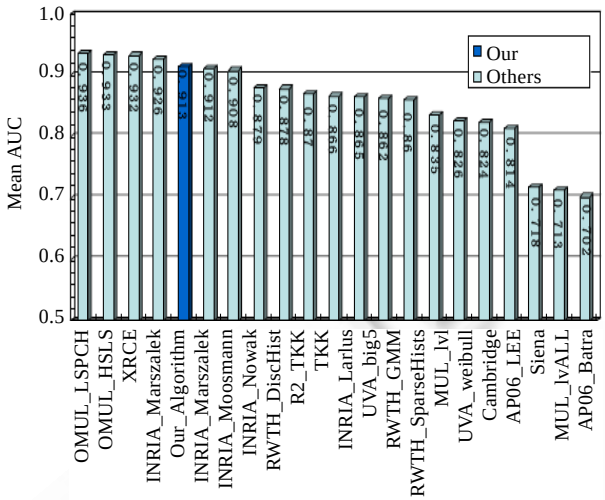


Fig.3 Comparison of mean AUC on PASCAL VOC 2006 dataset

图 3 PASCAL VOC 2006 数据集平均 AUC 比较

4 结 论

本文提出了基于视觉词统计建模与判别学习的视觉词软直方图图像表示方法,包括基于分类的软直方图方法和完全软直方图方法,用于解决物体识别问题.本文方法用高斯混合模型对视觉词进行建模,提高了局部特征与视觉词间相似度度量的准确性;利用判别学习方法估计视觉词模型中的未知参数,增强了视觉词统计模型对于局部特征的区分能力;在局部特征与视觉词相似度度量的基础上,获得了视觉词软直方图图像表示方法,降低了局部特征匹配的混淆误差.所提出的图像表示方法在两个常用的物体识别图像数据集上进行了实验验证,实验结果是:所提出的软直方图表示优于视觉词袋硬直方图表示.同时,相对于目前常用的高斯模型,采用高斯混合模型建模视觉词,能够提高物体识别正确率.此外,所采用的判别学习算法能够改善生成学习算法的学习结果.这些实验结果以及与同类算法的比较表明,本文算法是有效的.

本文的后续工作包括将类特定和空间位置信息引入视觉词的软直方图,以进一步提高图像局部特征表示的区分能力.同时,将在更多计算机视觉应用问题中验证所提出的图像表示方法,如图像标注与检索、物体定位、视频监控等.

References:

[1] Csurka G, Dance C, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: Proc. of the Workshop on Statistical Learning in Computer Vision in ECCV. Prague: Springer-Verlag, 2004.

[2] Zhang J, Marszalek M, Lazebnik S, Schmid C. Local features and kernels for classification of texture and object categories: A comprehensive study. Int'l Journal of Computer Vision, 2007,73(2):213−238. [doi: 10.1007/s11263-006-9794-4]

[3] Han DF, Li WH, Guo W. Object classification based on latent local spatial relations learning. Chinese Journal of Computers, 2007, 30(8):1286−1294 (in Chinese with English abstract).

[4] Joachims T. Text categorization with support vector machines: Learning with many relevant features. In: Proc. of the 10th European Conf. on Machine Learning (ECML'98). Chemnitz: Springer-Verlag, 1998. 137−142.

[5] Escamilla-Ambrosio PJ, Lieven N. Soft-Histogram degradation analysis of a tie bar of a rotor-head structure. Journal of Aircraft, 2008,45(6):2161−2164. [doi: 10.2514/1.34214]

[6] Farquhar J, Szedmak S, Meng H, Shawe-Taylor J. Improving “bag-of-keypoints” image categorization: Generative models and PDF-kernels. Technical Report, University of Southampton, 2005. http://eprints.pascal-network.org/archive/00008157/01/Improving_bag-of-keypoints_image_categorisation_Generative_Models_and_PDF-Kernels.pdf

[7] Perronnin F. Universal and adapted vocabularies for generic visual categorization. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008,30(7):1243−1256. [doi: 10.1109/TPAMI.2007.70755]

[8] Winn J, Criminisi A, Minka T. Object categorization by learned universal visual dictionary. In: Proc. of the ICCV. Beijing: IEEE Computer Society Press, 2005. 1800−1807. [doi: 10.1109/ICCV.2005.171]

[9] Yang L, Jin R, Sukthankar R, Jurie F. Unifying discriminative visual codebook generation with classifier training for object category recognition. In: Proc. of the CVPR. Anchorage: IEEE Computer Society Press, 2008. 1−8.

[10] Carneiro G, Chan AB, Moreno PJ, Vasconcelos N. Supervised learning of semantic classes for image annotation and retrieval. IEEE Trans. on Pattern Analysis and Machine intelligence, 2007,29(3):394−410. [doi: 10.1109/TPAMI.2007.61]

[11] Liu XB, Jia YD, Chen XF, Deng Y, Fu H. Image classification using the max-min posterior pseudo-probabilities method. Technical Report, BIT-CS-20080001, Beijing: Beijing Institute of Technology, 2008. http://isc.cs.bit.edu.cn/faculties/liuxiabi/papers/2008_1.PDF

[12] Hansen MH, Yu B. Model selection and the principle of minimum description length. Journal of American Statistical Association, 2001,96(454):746−774. [doi: 10.1198/016214501753168398]

[13] Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 1977,39(1):1−38.

[14] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning. In: Proc. of the CVPR. Madison: IEEE Computer Society Press, 2003. 264−271.

[15] Everingham M, Zisserman A, Williams CKI, Van Gool L. The PASCAL visual object classes challenge 2006 (VOC2006) results. 2006. http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf

[16] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. Int’l Journal of Computer Vision, 2004,60(1):63−86. [doi: 10.1023/B:VISI.0000027790.02288.f2]

[17] Lowe DG. Distinctive image features from scale-invariant keypoints. Int’l Journal of Computer Vision, 2004,60(2):91−110. [doi: 10.1023/B:VISI.0000029664.99615.94]

[18] Sivic J, Russell BC, Efros AA, Zisserman A, Freeman WT. Discovering objects and their location in images. In: Proc. of the ICCV. Beijing: IEEE Computer Society Press, 2005. 370−377. [doi: 10.1109/ICCV.2005.77]

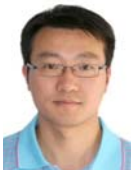
[19] Opelt A, Fussengger M, Pinz A, Auer P. Generic object recognition with boosting. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(3):416−431. [doi: 10.1109/TPAMI.2006.54]

[20] Kapoor A, Grauman K, Urtasun R, Darrell T. Active learning with Gaussian processes for object categorization. In: Proc. of the ICCV. Rio de Janeiro: IEEE Computer Society Press, 2007. 1−8.

附中文参考文献:

[3] 韩东峰,李文辉,郭武.基于潜在局部区域空间关系学习的物体分类算法.计算机学报,2007,30(8):1286−1294.

王彦杰(1982－),男,山西太原人,博士,工程师,主要研究领域为计算机视觉,模式识别.



刘峡壁(1972－),男,博士,副教授,博士生导师,CCF 会员,主要研究领域为模式识别,计算机视觉,机器学习,信息检索.

贾云得(1962－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机视觉,模式识别,智能系统.