基于多信息融合的视觉目标类识别算法研究

江爱文 王春恒 肖柏华 程 刚
(中科院自动化研究所复杂系统与智能科学重点实验室 北京 100190)

摘 要 视觉目标类识别是计算机视觉研究领域中的最具挑战性的难题之一,目前仍有许多问题没有得到很好的解决。近年来提出的空域金字塔直方图特征表示,在描述特征点集分布属性方面取得了比较好的实验效果。但是由于其描述的信息不全面,在性能上仍有较大改进余地。从信息互补性角度出发,提出了基于多信息融合的集成策略,将空域金字塔直方图表示与费舍分数表示各自描述的优势相结合,用于视觉目标类识别。实验证明该策略是有效的,在所进行测试的所有类别上相比单信息识别的性能均取得了一致性提高。

关键词 视觉目标类识别,空域金字塔直方图,费舍分数表示,多信息融合
中图法分类号 TP391 **文献标识码** A

Multi-information for Visual Object Categorization

JIANG Ai-wen WANG Chun-heng XIAO Bai-hua CHENG Gang
(Key Laboratory of Complex System and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

Abstract Visual object categorization(VOC) is one of the most difficult challenges in computer vision. Spatial pyramid histogram has been proposedd in recent years as an effective way to deal with features sets. However, there remains a large space for improvement. We made use of the respective advantage of spatial pyramid histogram and fisher score representation and proposed to use multi-information for recognition from information complement point view. The experiment results confirm our strategy, and our proposed algorithm consistently boosts the performance of all classes compared with their respective performances.

Keywords Visual object categorization, Spatial pyramid histogram, Fisher score representation, Multi-information combination

1 引言

在计算机视觉研究领域中视觉目标类识别是最具挑战性的难题之一。它不仅需要克服传统的特定目标识别问题中碰到的背景、视角等变化,同时还需要解决类间差异与类内差异的矛盾,难度更大,更具挑战性。自2005年起,每年一次轮流在ICCV,ECCV顶级会议上举行PASCAL-VOC竞赛,同时在CVPR会议上与之相关的论文也占了相当的比例,吸引了来自世界各国相关研究小组的广泛兴趣。虽然取得长足的进步,但仍然有许多问题没有得到很好的解决。其中目标物图像的特征表示是影响一个识别系统性能的重要因素。近年来,研究者提出了许多特征建模方法,具有代表性的有:(1)基于外观的模型(appearance-based model),如词袋模型(Bag of words)[1,2];(2)星座模型(constellation model)[3],用于描述目标物的形状;(3)描述物体的轮廓信息[4]。这些模型各有优缺点。但目前主流的并且行之有效的方法仍是基于外观的模型,具有代表性的是词袋(Bag of words)表示方法。

在基于外观的模型中,图像一般采用局部描述算子进行表示[5]。与传统理解的图像特征抽取不同,一副图像并不直接表示成一个固定长度的特征向量,而是被表示为由大量具有尺度不变性的局部描述算子组成的特征点集。因此,如何有效组织这些特征点集,是一个需要研究的课题。

Grauman与Darrell提出采用金字塔匹配核(pyramid match kernel)的方法[1]。他们按金字塔形式将特征空间划分为不同粒度级别(pyramid resolution),然后将每个粒度级别上的特征匹配数目进行加权求和,作为最后两两特征点集之间的相似度。Lazebnik等人提出的空域金字塔匹配策略可以认为是与文献[1]相"正交"的方法[2]。他们不是在特征空间上,而是直接在原图像上将图像像素空间划分成不同层次的栅格,然后每层次进行特征匹配,加权求和作为两幅图像的相似度。空域金字塔匹配策略在一定程度上考虑了图像内容的空间分布信息。

词袋(Bag of words)特征表示是最具代表性的一种方法。它的思想源自文本分类,类比图像与文本,将图像看成由大量的"视觉单词"(visual words)组成。这些"视觉单词"对应于图像的局部区域。首先将这些"视觉单词"聚类到预先学习得

到稿日期:2009-10-30 返修日期:2010-01-15 本文受国家自然科学基金(No. 60802055与No. 60835001)资助。
江爱文(1984—),男,博士生,主要研究方向为图像处理、模式识别,E-mail:aiwen.jiang@ia.ac.cn;**王春恒**(1972—),男,博士生导师,主要研究方向为图像处理、模式识别、计算机视觉等;**肖柏华**(1974—),男,硕士生导师,主要研究方向为图像处理、模式识别、计算机视觉等;**程 刚**(1982—),男,博士生,主要研究方向为图像处理、模式识别等。

到的词典(codebook),然后图像被表示成"词频(TF)"直方图向量。因此大部分的工作都集中在如何通过聚类来估计"视觉词典"(visual codebook)。主要的方法可以归为两类:(1)"硬"聚类算法,如K均值聚类算法;(2)"软"聚类算法,如高斯混合模型[6]。

词袋表示的优点是简单、有效,能够将不等数目的特征点集映射到固定长度的特征向量。但是它仅仅是无序特征点集的一种组织方式,信息描述不全面。本文从信息互补性的角度出发,提出多信息融合的集成策略,将空域金字塔式的直方图特征与费舍分数特征表示相结合,用于视觉目标类识别。

引入费舍分数表示(fisher score representation)是为了有效结合判别模型与生成模型的优点,形成一个融合生成模型-判别模型的混合系统。Florent Perronnin 等人将费舍分数表示用于高斯混合模型[7]。Alex D. Holub 则把费舍分数用在星座模型上[8]。费舍分数表示之所以受到重视,是因为它在生成模型的梯度空间上反映了特征的生成过程。生成模型的Log似然函数(Log-likelihood)参数对应的梯度值,能够描述该参数在样本生成的过程中是如何起作用的。因此log似然函数的梯度空间能够反映生成模型对于生成过程所做的结构性假设[9]。我们有理由相信,费舍分数表示的信息与词袋词频直方图表示的信息具有一定互补性,二者相结合有助于提高目标类识别的性能。

实验表明,我们的策略是有效的,在测试的所有目标类别上,相比单信息识别的性能均有所提高。

在接下来的篇幅中,将在第2节具体描述我们的信息互补集成策略,包括空域金字塔直方图的构建、费舍分数表示以及它们的融合;在第3节将描述我们的实验细节及在PASCAL VOC2007数据集上的实验结果;最后给出结论。

2 基于多信息融合的集成策略

2.1 空域金字塔直方图

针对空域直方图特征表示,主要延续了文献[2]的策略。为了完备性,简要介绍一下主要的流程。流程图如图1所示。

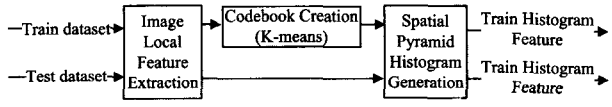

图1 构建空域金字塔直方图流程图

从样本图像中抽取大量的局部描述特征,采用聚类算法聚类,得到视觉词典。同时将图像按空间均匀划分成1×1,2×2两种,如图2所示。每种划分中的每一个方格(grid)均用词袋词频直方图形式表示。最后所有的直方图向量接合,组成图像最终的空域金字塔式直方图特征。

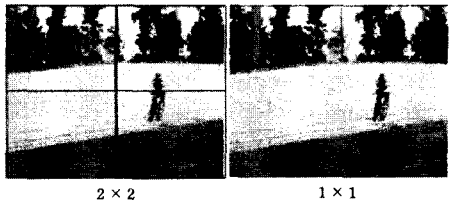

图2 二级空域划分示意图

这里不同之处在于,文献[2]采用了密集型采样描述,本文采用的是具有尺度不变性的局部感兴趣区域SIFT描述[11]。视觉词典用K均值聚类算法得到。

2.2 费舍分数表示(Fisher score representation)

费舍分数表示采用的是用生成模型估计特征点集的分布。令 $X=\{x_i, i=1,\cdots,N\}$ 表示特征点集。$\lambda=\{w_i,\mu_i,\Sigma_i, i=1,\cdots,M\}$ 表示高斯混合模型参数,其中 w_i,μ_i,Σ_i 分别表示第 i 个高斯成分对应的权重、均值向量以及协方差矩阵。N 表示特征点集的大小,M 表示高斯混合模型的高斯成分个数。

无序特征点集采用高斯混合模型(Gaussian Mixture-Model)$p(x_i|\lambda)$ 来生成:

$$p(x_i|\lambda)=\sum_{k=1}^{M} w_k p_k(x_i|\lambda)$$

式中,$\sum_{k=1}^{M} w_k=1$,$p_k(x|\lambda)=(2\pi)^{-D/2}|\Sigma_k|^{-1}\exp(-\frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k))$。

对参数 λ 的估计采用最大似然函数估计的方法,具体采用EM算法进行迭代估计。

$$\text{E-step}:\gamma_i(k)=\frac{w_k p_k(x_i|\lambda)}{\sum_{j=1}^{M} w_j p_j(x_i|\lambda)} \tag{1}$$

$$\text{M-step}:w_k=\frac{\sum_{i=1}^{N}\gamma_i(k)}{N} \tag{2}$$

假设协方差矩阵为对角阵,即 $\sigma_k^2=diag(\Sigma_k)$。这样的假设是合理的,因为任何分布都可以近似地由若干个对角协方差高斯函数加权累加得到。简单的对角阵假设可以大大减少带估计参数数目,可以有比较好的泛化能力,同时降低运算成本。

采用的费舍分数表示是定义在Log似然函数的梯度空间上,具体定义如下:

$$\phi(X,\lambda)=\nabla_\lambda L(X|\lambda)=\nabla_\lambda\log(p(X|\lambda))$$

Log似然函数(Log-likelihood)参数对应的梯度值,能够描述该参数在样本生成的过程中是如何起作用的。因此log似然函数的梯度空间能够反映生成模型对于生成过程所做的结构性假设[9]。

给定一个特征集合 $X=\{x_t,t=1,\cdots,T\}$,采用式(3)和式(4)将特征点集合映射到费舍分数(fisher-score)空间上:

$$\frac{\partial L(X|\lambda)}{\partial\mu_k^d}=\sum_{t=1}^{T}\gamma_t(k)\left(\frac{x_t^d-\mu_k^d}{(\sigma_k^d)^2}\right) \tag{3}$$

$$\frac{\partial L(X|\lambda)}{\partial\sigma_k^d}=\sum_{t=1}^{T}\gamma_t(k)\left(\frac{(x_t^d-\mu_k^d)^2}{(\sigma_k^d)^3}-\frac{1}{\sigma_k^d}\right) \tag{4}$$

对梯度向量进行归一化,得到

$$f_{\mu_k^d}^{1/2}*\partial L(X|\lambda)/\partial\mu_k^d$$

$$f_{\sigma_k^d}^{1/2}*\partial L(X|\lambda)/\partial\sigma_k^d$$

式中,$f_{\mu_k^d}=\frac{T*w_k}{(\sigma_k^d)^2}$,$f_{\sigma_k^d}=\frac{2*T*w_k}{(\sigma_k^d)^2}$。

最终图像在fisher-score空间上得到的特征向量可以表示成以这些归一化的偏微分算子值为元素的特征向量。

2.3 多信息融合策略

空域金字塔式直方图表示和费舍分数表示可以认为是对无序、不等量的特征点集合的两种不同的组织方式。它们从不同的角度描述了点集的分布属性,具有一定信息互补性。我们采取两种简单的融合策略。

2.3.1 线性加权融合

核函数级别上进行融合的策略。将空域金字塔直方图表

示的核函数与费舍分数表示的核函数进行加权累加融合，得到一个包含多种信息的核函数：

$$K(x,y)=\alpha * K_{hist}(x,y)+(1-\alpha) * K_{fisher}(x,y) \quad (5)$$

式中，α 取值范围为[0,1]。在本文中，α 取经验值 0.5。

2.3.2 两级分类器融合

分类器级别上进行融合的策略。第一级分类器的输出是概率值(class score)，表示属于某一类的概率。然后以这些概率分数作为下一层分类器的输入向量，得到最终的分类结果(output score)。具体如图 3 所示。

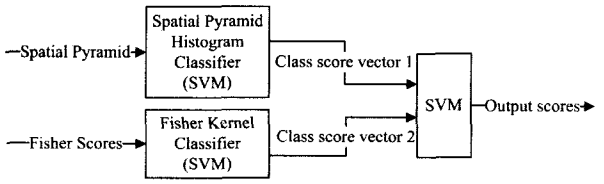

图 3 二级分类器融合示意图

本文中所有分类器均采用基于核函数的分类器。对于空域金字塔直方图表示，采用的核函数是扩展的高斯核函数：$K(x_i,x_j)=\exp(-\sum_{ch}(\omega_{ch} * D(x_i,x_j))$

距离函数 D 采用卡方距离：

$$D(x_i,x_j)=\sum_k \frac{(x_{i,k}-x_{j,k})^2}{x_{i,k}+x_{j,k}}$$

对于费舍分数表示采用的则是线性核，即费舍分数表示的内积形式。

3 实验设计与分析

密集型描述算子(dense grid descriptors)和稀疏型描述算子(sparse descriptors)是目标类识别算法中广泛采用的两种形式。对于密集型的描述，局部区域的尺度选择具有很大的主观性。稀疏型描述基于局部检测子检测到的感兴趣区域，严格按照尺度空间理论，具有尺度不变的特性，对于图像目标物尺度变换比较大的特点有着较好的自适应性。因此在本文实验中，我们将采用由局部检测子(Hessian Laplacian 检测子[10]和 Difference-of-Gaussian 检测子[11])组成的 SIFT 描述表示。

实验采用的数据集是 PASCAL VOC2007 竞赛数据集[12]。数据集中总共包含了 20 个目标大类，分别是人、鸟、猫、狗、牛、马、羊、飞机、自行车、船、巴士、轿车、摩托车、火车、瓶子、椅子、餐桌、盆栽、沙发和显示器，共 5011 幅训练图像和 4952 幅测试图像。部分类别样本如图 4 所示。

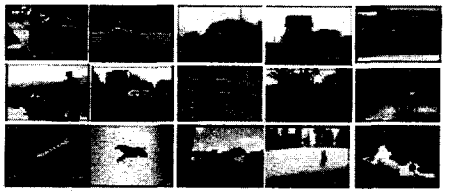

图 4 PASCAL VOC2007 数据集样本图片

3.1 空域金字塔直方图表示

从训练集中每类随机抽取 20 幅图像，共计 400 幅图像，作为“视觉词典”构建的样本。

利用数据的标定信息，将检测子检测得到的局部特征描述划分为两部分：一部分分布在标定目标类区域内，作为该类“前景”局部区域特征；另一部分分布在标定区域外，作为“背景”类特征。然后，对于每类的“前景”特征，采用聚类算法得到 200 个“视觉单词”；对于“背景”特征，通过聚类得到 1000 个左右的“视觉单词”。因此“视觉词典”由 200 * 20+1000=5000 个视觉单词组成。

图像被空间划分成 1×1 和 2×2 两种。每种划分中的局部特征表示成直方图特征，计算卡方距离。最终构建扩展高斯核函数。

3.2 费舍线性核(Fisher Kernel)

在费舍分数表示的生成模型训练过程中，同样每类抽取部分样本，将检测得到的局部描述特征集用于高斯混合模型的训练。为了在一定程度上避免过拟合，SIFT 特征采用 PCA 主成分分析方法从 128 维降至 50 维。高斯混合模型中高斯个数选为 200 个左右。

费舍线性核由费舍分数表示(fisher score representation)向量内积而成。

3.3 非对称 Bagging 策略(Asymmetric Bagging Strategy)

PASCAL VOC2007 数据存在较为严重的数据不均衡和小样本问题。为了减轻这些问题对性能带来的影响，我们在实验过程中采用非对称的 bagging 策略[13]。

所谓非对称 Bagging 策略思想是只在负面样本上执行 bootstrap 采样，正面样本全部利用。原因在于对于每一目标类来说，负面样本数目要远远多于正面样本。

对于每一种特征表示方式，采用 bootstrap 对训练集进行采样，训练多个分类器。bootstrap 训练得到的多个分类器采用 sum rule [14]的方式聚合。在试验中，我们试验了不同次数的采样，发现对于本实验 bagging 次数超过 5 次，性能提升不会太大。因此在本实验中，我们对每种特征表示方式进行 5 次非对称 bagging 策略。

3.4 实验结果

所有的分类实验均采用二分类形式。即对于某一目标类，所有包含该类的图片归为正面样本，否则为负面样本。然后计算该类的召回率和准确率。

我们采用 PASCAL VOC2007 测评中标准采用的平均准确率(Average Precision) 作为我们实验的每类分类性能的衡量标准：

$$AveP=\frac{\sum_{r=1}^{N}(P(r)\times rel(r))}{N}$$

式中，$rel(r)$ 表示某一次排序的相关度(0 或 1)，$P(r)$表示相应排序的准确率，N 表示进行排序的次数。

具体操作是：对于每一类分类结果，用属于某一类目标物的概率值表示。将(0,1)区间 N 等分($bi=1/N,2/N,\cdots,k/N,\cdots,1$)，作为 N 次排序检索的召回率下界。每一次选择召回率 $recall\ rate \geqslant bi$ 时，最优的准确率为 $P(r)$。重复 N 次，计算平均召回率。

MAP(mean average precision)用于衡量最终的全类别上的分类性能：

$$MAP=\sum_{i=1}^{num\_class} AveP(i)$$

为了公正性，我们采用 VOC2007 提供的标准工具包来评价分类准确率[12]。实验中采用的分类器是 LibSVM[15]。实验结果如表 1 所列。

表 1 各类别的平均准确率

Class	Spatial histogram	Fisher scores	Linear Weighted Combination	Two Level Combination
Aero-plane	0.630	0.610	0.648	0.663
bicycle	0.522	0.500	0.543	0.563
bird	0.384	0.374	0.426	0.426
boat	0.510	0.490	0.530	0.530
bottle	0.247	0.231	0.265	0.275
bus	0.460	0.420	0.450	0.481
car	0.650	0.595	0.670	0.686
cat	0.470	0.450	0.482	0.511
chair	0.455	0.417	0.468	0.485
cow	0.289	0.214	0.312	0.336
dining table	0.334	0.300	0.337	0.359
dog	0.360	0.335	0.387	0.404
horse	0.615	0.578	0.638	0.669
motorbike	0.520	0.501	0.559	0.556
person	0.780	0.779	0.800	0.810
Potted plant	0.217	0.197	0.235	0.247
sheep	0.315	0.298	0.324	0.335
sofa	0.375	0.350	0.377	0.400
train	0.620	0.562	0.624	0.635
TV/monitor	0.410	0.387	0.422	0.438
Mean Average Precision	0.458	0.429	0.475	0.490

我们根据信息的互补性，进行两种简单的融合策略。值得注意的是，为了验证这两种表示的信息互补性，我们相比参赛队伍的算法中用到的特征而言，采用了比较简单和单一的局部特征，但我们最终的分类性能却接近了 VOC2007 竞赛的前三水平（VOC 第三名水平（*mean average precision* = 0.503）。从表 1 的实验结果可以看出，本文提出的两种信息融合策略在所有的目标类上相比单种信息均取得了一致提高。这说明空域金字塔直方图表示和费舍分数表示确实存在一定的信息互补性，我们的策略是行之有效的。这两种表示具有一定普遍意义，可以同时用于多种特征上，因此相信我们融合的策略运用到多个特征上时，性能又会有不同程度的提高。

结束语 视觉目标类识别是当今计算机视觉研究领域的重大挑战之一，吸引着世界范围内相关研究者的广泛兴趣。近年来提出的空域金字塔直方图表示在实际的研究中就如何有效地组织无序、不等数量的特征点集，取得了比较好的效果。但由于其信息描述不全面，性能上存在着一定的瓶颈。本文从信息互补性角度出发，提出多种信息融合的集成策略，结合空域直方图特征与费舍分数表示二者的优点用于视觉目标类识别。实验结果表明，我们的策略是行之有效的。通过信息融合，在性能上相比单信息，在所有目标类别上均取得了一致性的提高。

参 考 文 献

[1] Grauman K, Darrell T. The pyramid match kernel: Discriminative classification with sets of image features[C]//ICCV. 2005

[2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]//CVPR. 2006

[3] Fergus R, Perona P, Zisserman A. Object class recognition using unsupervised scale-invariant learning[C]//CVPR. 2003

[4] Ferrari V, Fevrier L, Jurie F, et al. Groups of adjacent contour segments for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(1): 36-51

[5] Mikolajczyk K, Leibe B, Schiele B. Local features for object class recogniton[C]//ICCV. 2005

[6] Farquhar J, Szedmak S, Meng H, et al. Improving "bag-of-keypoints" image categorization: generative models and pdf-kernels [R]. University of Southampton, 2005

[7] Perronnin F, Dance C. Fisher kernel on visual vocabulries for image categorization[C]//CVPR. 2007

[8] Holub A D, Welling M, Perona P. Hybrid generative-discriminative visual categorization[J]. Internal Journal Computer Vision, 2008, 77(1): 239-258

[9] Gales M, Layton M. Maximum margin training of generative kernels[R]. University of Cambridge, 2004

[10] Mikolajczyk K, Schmid C. Scale and affine invariant interest point detectors[J]. Internal Journal Computer Vision 2004, 60(1): 63-86

[11] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. Internal Journal Computer Vision, 2004, 60(2): 91-110

[12] The PASCAL Visual Object Classes Challenge[EB/OL]. available on http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/index.html, 2007

[13] Tao D, Tang X, Li X, et al. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(7): 1088-1099

[14] Kittler J, Hatef M, Duin P W, et al. On combining classifiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(3): 226-239

[15] Chang C C, Lin C J. LIBSVM: a library for support vector machines[EB/OL]. Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001

(上接第 244 页)

这些结论使人们对于规律挖掘的阶梯型、层次（渐进）性有了更进一步的认识，它在规律挖掘中具有更重要的作用。

参 考 文 献

[1] Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Sciences, 1982, 11: 341-356

[2] Shi Kai-quan. S-rough sets and its application in diagnosis-recognition for disease[J]. IEEE Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002, 1: 50-54

[3] Shi Kai-quan, Cui Yu-quan. *F*-decomposition and $\bar{F}$-reduction of S-rough sets[J]. An International Journal Advances in Systems Sciences and Applications, 2004, 4: 487-499

[4] Shi Kai-quan, Chang Ting-cheng. One direction S-rough Sets [J]. International Journal of Fuzzy Mathematics, 2005, 2: 319-334

[5] Shi Kai-Quan. Two direction S-rough Sets [J]. International Journal of Fuzzy Mathematics, 2005, 2: 335-349

[6] Zhang Guan-Yu, Du Ying-Ling, Qiu Yu-Feng. Knowledge Law and Attribute Disturbance of Law[J]. 数学季刊, 2008, 23(2): 245-251

[7] Zhang Guan-yu, Du Ying-ling. Attribute Disturbance of Knowledge and Attribute Disturbance Theorems[J]. 数学季刊, 2008, 23(4): 574-581

[8] Yin Shou-feng, Shi Kai-quan, Hu Hai-qing. Two direction S-rough extension communication and its heredity-variation characteristics[J]. International Journal of Fuzzy Mathematics, 2006, 4: 408-413

[9] 史开泉，崔玉泉. S-粗集与粗决策[M]. 北京：科学出版社，2006：5-8

[10] Shi Kai-quan. Function S-rough sets and function transfer[J]. An International Journal Advances in systems Science and Applications, 2005, 1: 1-8

[11] 史开泉，姚炳学. 函数 S-粗集与规律辨识[J]. 中国科学(E)，2008, 4: 553-564

[12] Shi Kai-quan, Yao Bing-xue. Function S-rough sets and law identification[J]. Science in China(F), 2008, 5: 499-510