基于潜在局部区域空间关系学习的物体分类算法

韩东峰 李文辉 郭 武

(吉林大学计算机科学与技术学院 长春 130012)

(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)

摘 要 提出了一种物体分类模型——潜在局部区域空间关系模型及实现算法.模型描述物体各部分间的潜在空间关系,将抗缩放和仿射变换的特征区域方法与模型相结合,采用变分期望值最大方法进行学习.与同类空间关系模型算法比较,该文模型算法具有以下优点:(1)为非监督式学习模型;(2)可抵抗几何变换;(3)模型为稠密模型;(4)模型描述的是一种潜在的空间关系,这种关系对物体具有更本质的描述.在标准测试图像库上的实验表明了该算法在抵抗平移、旋转、尺度缩放、仿射变换和部分遮挡等难点问题上具有优势.

关键词 物体分类;潜在局部空间关系;图模型;变分期望值最大化;局部特征点

中图法分类号 TP391

Object Classification Based on Latent Local Spatial Relations Learning

HAN Dong-Feng LI Wen-Hui GUO Wu

(*College of Computer Science and Technology, Jilin University, Changchun* 130012)

(*Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun* 130012)

Abstract Latent Local Spatial Relations (LLSR) model is presented as a novel technique of learning spatial models for visual object classification. Combined the latent local spatial relations model with statistical visual words and variational expectation maximization, LLSR is developed as an implementation of object classification algorithm. LLSR uses an unsupervised process that can capture both spatial relations and visual words appearances simultaneously. In contrast to other methods which explicitly give some parameterized spatial models, the proposed algorithm uses a latent class model to reveal some certain latent spatial relations. The advantages of the proposed model include: (1) it uses an unsupervised learning paradigm which can avoid some manual controls; (2) it can resist some geometry transforms; (3) it is a dense model; (4) the spatial relations are latent which have more insight into describing the object structure. The experiments are demonstrated on some standard databases and show that LLSR is a promising model for solving object classification problems, especially for translation, rotation, scale, affine and part of occlusion.

Keywords object classification; latent local spatial relations; graph model; variational expectation maximization; local interest points

---

收稿日期:2007-03-05;修改稿收到日期:2007-05-24.本课题得到国家自然科学基金(50338030,60573182)、教育部博士点基金(20060183042)、吉林省科技发展计划项目基金(20040531,20060527)资助.**韩东峰**,男,1979年生,博士研究生,主要研究方向为计算机视觉、机器学习和模式识别.E-mail:handongfeng@gmail.com.**李文辉**,男,1961年生,博士,博士生导师,主要研究领域为计算机视觉和虚拟现实.**郭 武**,女,1972年生,博士研究生,主要研究方向为计算机图形学和虚拟现实.

1 引　言

物体分类识别是计算机视觉中最重要的研究课题之一.它通过对给定图像信息的提取与处理得到模型从而对其它的图像进行推断,内容包括是否包含某物体、物体的定位等.准确快速地建立模型并进行有效的分类识别具有广泛的用途.然而,以下几个问题始终是人们关心的和比较难以解决的:

(1)建模.怎样对分类识别问题进行建模;

(2)学习.怎样对模型学习;

(3)测试.怎样抵抗几何变换(平移、旋转、缩放、仿射)、遮挡等问题的影响.

目前,用于物体分类识别的方法可归纳为两种模型.一种是仅仅体现了物体的外部信息(特征区域、纹理、轮廓等)而没有给出构成物体的元素间的空间关系[1-4].该方法主要是使用统计信息来进行分类识别.另一种则既体现了物体的外在信息又包含元素间的空间关系(通常把物体分为一些组)[5-10].我们统称第一种方法为 bag-of-word 方法,后一种为 part-based-methods,显然后者具有更多的优点.近年来,研究者们提出了很多较好的 part-based-methods. Burl 等人[5]提出了联合高斯模型,能捕捉到所有组成物体的部分之间的空间关系. Weber 等人的方法[6]也是学习物体的各个组成部分并计算它们的联合分布概率. Fergus 等人[7]改进了这种模型,估计物体空间的以及外在的联合分布概率.该方法可以抵抗尺度变化,但是这种模型中的检测和定位算法多为启发式,而且物体一般由较少的部分组成. Agarwal 等人[8]把物体分成较多部分,并在学习物体空间关系时使用了一种特征有效的分类器,但是该方法需要很多的训练图像.还有一种应用比较多的模型是树形结构[9],它能有效地检测和定位一些类别的物体,如人体、人脸等.在这种模型中物体各个部分的关系构成一个树,但是在对图像的仿射变换(affine)下不能达到很好的检测效果. Crandall 等人[10]提出了一种介于这两种模型之间的 k-fans 模型,即用 k 个物体的组成部分与其它组成部分的关系作为物体的空间关系.若需模型简单易实现,可以把 k 值取小一些,若需模型相对稳定精确,则把 k 值取大一些.但该方法缺点是取 $k=2$ 时计算复杂性已相当高了.

然而,从实际分类问题的角度上讲,上述两类方法都不能同时兼顾几何变换(主要以仿射变换为难点)、遮挡等难点.本文提出了一种潜在局部区域空间关系模型,将物体的空间关系表示为一种概率的潜在关系,同时在实现中通过提取局部不变的特征点方法加入图像特征的统计信息.理论分析和实验表明,本文算法具有抵抗几何变换、遮挡等优点.

本文第 2 节介绍潜在局部区域空间关系模型及模型的学习过程;第 3 节对模型相应问题进行了讨论;第 4 节中给出具体实现细节;实验及分析对比在第 5 节中给出;最后,在第 6 节中进行总结并说明下一步的工作.

2 潜在局部区域空间关系

2.1 基本定义

在传统的文本检索中,所有文章由来自一个公共字典的单词组成,不同的文章组成文章集合.类似地,我们规定所有图像由一个共同、固定的特征码书组成,若干图像组成图像集合.包含同一物体或者同一场景的图像在特征码书的概率分布以及相互间的关系上存在一种潜在关系,这就是本文算法的基本出发点.为方便后续的讨论,同时结合传统文本检索的习惯,本文给出以下定义.

定义 1. 特征单词(word)是组成图像的最基本单位,特征单词在特征码书中以 $\{1,2,\cdots,M\}$ 索引.特征单词的提取以及码书由不同的问题规模决定,后续 4.1 节会有详细讨论.

定义 2. 特征区域由图像某些特征所决定,特征区域表示为 $r=\{w_1,w_2,\cdots,w_M\}$.一幅图像由特征区域组成,特征单词决定了特征区域,同时区域间的特征单词关系还决定了区域间的空间关系.

定义 3. 一幅图像(document)由 N 个特征区域组成,表示为 $d=\{r_1,r_2,\cdots,r_N\}$.

定义 4. 每幅图像包含 K 个主题,表示为集合 $Z=\{z_1,z_2,\cdots,z_K\}$.主题往往是未知的,也是人们感兴趣的,在机器学习中通常表示为隐形变量.

定义 5. V 幅图像可以组成一组图像数据集合,表示为 $D=\{d_1,d_2,\cdots,d_V\}$.

2.2 潜在局部区域空间关系模型(Latent Local Spatial Relations Model)

潜在局部区域空间关系模型的核心在于引入特征区域,图像是由区域组成的,并且强调区域间的关系.而每个特征区域是由特征单词所组成(这区别于

其它方法,如文献[1-2]等,这些方法认为图像直接由特征单词组成,从而没有考虑空间关系).潜在局部区域空间关系模型(LLSR)如图1所示,其中阴影变量代表可观测变量.

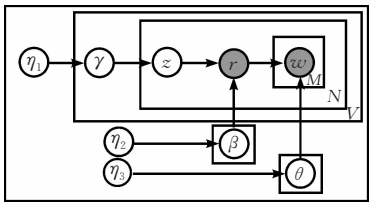


图1 潜在局部空间关系模型的图模型

LLSR包括下面几个概率生成步骤:

(1)对于主题变量 z,z 是一个 K 维的单位向量,K 为主题类别数目. $z_k=1$ 代表第 k 个主题被选择. z_k 服从参数为 γ 的多项式分布,$P(z_k/\gamma)\sim Mult(\gamma)$.

(2)变量 γ 是多项式分布的参数,γ 服从多项式分布的先验对偶分布——狄氏分布. K 维狄氏分布具有性质 $\sum_{i=1}^{K}\gamma_i=1$,概率分布形式如下,

$$P(\gamma/\eta_1)\sim Dir(\eta_1)=\frac{\Gamma\left(\sum_{i=1}^{K}\eta_{1i}\right)}{\prod_{i=1}^{K}\Gamma(\eta_{1i})}\gamma_1^{(\eta_{11}-1)}\cdots\gamma_K^{(\eta_{1K}-1)},$$

其中,$\Gamma(\cdot)$ 为 Gamma 函数.

(3)图像由特征区域组成,每个特征区域属于某个主题变量 z_k. 特征区域 r_k 服从参数为 β 的条件多项式分布,$r_k\sim P(r_k/z_k,\beta)$.

(4)每个特征区域由特征单词所决定,特征单词 w_t 服从参数为 θ 的条件多项式分布,$w_t\sim P(w_t/r,\theta)$.

(5)对应于参数 β 和 θ 也有相应的先验分布 η_2,η_3,$P(\beta/\eta_2)\sim Dir(\eta_2)$,$P(\theta/\eta_3)\sim Dir(\eta_3)$.

LLSR是一个层次图模型. 狄氏分布参数是类别层次,每产生一次对应于产生某一类图像. 多项式分布参数 γ 对应于图像层次,每幅图像对应一个 γ. 隐形主题变量属于区域层次,每个区域对应一个主题. 最后是特征区域和特征单词,每个区域由不同的特征单词所组成. 变量 r 服从独立同分布(iid),来源于某个主题 z_k. 上述的LLSR模型中,γ,β 和 θ 的分布为狄氏分布,由于狄氏分布是多项式分布的先验对偶概率分布,这样的选择是为了后验概率推理的方便[11-13]. LLSR模型主要的特点在于加入变量 r 对物体的空间关系进行描述. 后续会看到,可以将这种区域间的关系通过特征单词间的关系联系起来,进而可以使用学习的方法来学习区域间的关系. 同时,这种关系是一种潜在关系,这也是LLSR的来由.

2.3 参数学习

LLSR模型是一个含有隐形变量的贝叶斯网络. 求解方法是期望值最大化方法(EM). EM方法包括E步骤和M步骤. 给定一组观测图像集合 $D=\{d_1,d_2,\cdots,d_V\}$,目标是找出使观测数据对数似然概率最大的参数,即

$$(\hat{\eta}_1,\hat{\eta}_2,\hat{\eta}_3)=\arg\max_{\eta_1,\eta_2,\eta_3}\log P(D/\eta_1,\eta_2,\eta_3) \tag{1}$$

$$L=\log P(D/\eta_1,\eta_2,\eta_3)=\sum_{i=1}^{V}\log P(d_i/\eta_1,\eta_2,\eta_3) \tag{2}$$

对数似然函数为 L,式(2)可以展开,得

$$\begin{aligned}L&=\log P(D/\eta_1,\eta_2,\eta_3)\\&=\log\iint P(D,\beta,\theta/\eta_1,\eta_2,\eta_3)\mathrm{d}\beta\mathrm{d}\theta\\&=\log\iint q(\beta,\theta)\frac{P(D,\beta,\theta/\eta_1,\eta_2,\eta_3)}{q(\beta,\theta)}\mathrm{d}\beta\mathrm{d}\theta\\&\geq\iint q(\beta,\theta)\log\frac{P(D,\beta,\theta/\eta_1,\eta_2,\eta_3)}{q(\beta,\theta)}\mathrm{d}\beta\mathrm{d}\theta\\&=\underbrace{E_q[\log P(D/\eta_1,\beta,\theta)]}_{L_A}+\underbrace{E_q[\log P(\beta,\theta/\eta_2,\eta_3)]}_{L_B}-\\&\quad\underbrace{E_q[\log q(\beta,\theta)]}_{L_C}\end{aligned} \tag{3}$$

其中,式(3)L_A 中的 $\log P(D/\eta_1,\beta,\theta)$ 类似 LDA[11-12] 模型,并且每部分都有较复杂的积分操作,因此一般使用近似求解方法. L_A 部分可以进一步写为

$$\log P(D/\eta_1,\beta,\theta)=l(\alpha,\phi;\eta_1,\beta,\theta)+KL(q(\gamma,z/\alpha,\phi)\parallel P(\gamma,z/D,\eta_1,\beta,\theta)) \tag{4}$$

其中,α,ϕ 为引入的变量. $KL(\cdot)$ 为相对熵. 式(4)表明最大化 $l(\alpha,\phi;\eta_1,\beta,\theta)$ 等同于最小化 $KL(q(\gamma,z/\alpha,\phi)\parallel P(\gamma,z/D,\eta_1,\beta,\theta))$. 由此首先应该估计 α,ϕ 的值,然后将其作为 $\log P(D/\eta_1,\beta,\theta)$ 的近似值. 进而可以继续估计 η_1,β,θ 的值. 该方法称为变分期望值最大法(Variational EM,VEM)[11-13]. 基于VEM算法的参数估计步骤如下.

对于训练图像,优化参数 α,ϕ,更新公式为

E步骤:

$$\alpha_i^{\mathrm{new}}=\eta_{1i}+\sum_{j=1}^{N}\phi_{ji} \tag{5}$$

$$\phi_{ji}^{\mathrm{new}}\propto\beta_{ir_j}\exp\left\{\psi(\gamma_i)-\psi\left(\sum_{t=1}^{K}\gamma_t\right)\right\} \tag{6}$$

其中,$\psi(\cdot)$ 为 Digamma 函数.

使用E步骤中得到的参数,求解使式(2)最大化的 β,θ,

η_1,η_2,η_3，各参数的更新公式为

M 步骤：

$$\beta_{ki}^{\text{new}} \propto \sum_{d=1}^{V}\sum_{t=1}^{N} \phi_{dtk}\, r_{dti} \tag{7}$$

$$\theta_{jl}^{\text{new}} \propto \sum_{d=1}^{V}\sum_{n=1}^{M} r_{dnj} w_{dnl} \tag{8}$$

$$\eta_i^{\text{new}} = \eta_i^{\text{old}} - \boldsymbol{H}(\eta_i^{\text{old}})^{-1}\boldsymbol{g}(\eta_i^{\text{old}}),\ \ i=1,2,3 \tag{9}$$

其中，$\boldsymbol{H}(x)$和 $\boldsymbol{g}(x)$为 x 处 Hessian 矩阵及相应的梯度矩阵.

3　问题讨论

潜在局部区域空间关系模型的核心思想是通过局部区域的相互关系对物体进行刻画. 在第 2 节中给出了其模型，本节对实现中的细节进行讨论.

3.1　区域拓扑结构以及区域间关系的表示

本文使用特征点方法，在一个局部区域内，多个特征点具有空间拓扑关系. 通过设计，这种关系既可以表示局部区域内的拓扑结构，还可以表示区域间的相互关系，并且这种相互关系是通过统计概率来表示的一种潜在关系. LLSR 模型应用于物体分类识别，可以表示为一个三维表，如图 2 所示，r,w,d 分别代表区域、特征单词和图像. 对于每一幅图像来说，该图像可以由多个区域表示，其中每个区域是由特征单词组成. 图中平行于 $r-w$ 的平面组成一幅图像，多幅图像构成整个数据集. 对于某一幅图像来讲，每一列代表一个区域，不同的特征单词决定该区域的拓扑结构.

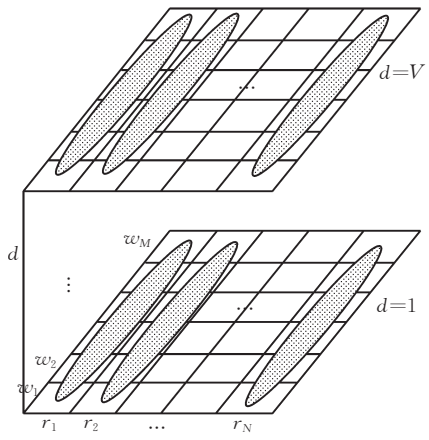


图 2　区域关系的三维表示

模型中如何确定区域是个比较重要的问题，本文将一个特征单词为中心引起的区域规定为一个局部区域，区域内有多个特征单词与之发生关系. 如图 3 所示，rw_{12}，rw_{15}代表这些特征单词间的关系. 这种关系同时也定义了该区域的拓扑结构并且可以抵抗几何变换. 由于使用的是特征单词引起的区域，因此区域间的特征单词之间的这种关系不仅反映了区域的空间拓扑结构，还可以反映区域与区域间的关系. 本文使用区域内，特征单词与引起该区域的特征单词的计数来表示这种区域拓扑结构.

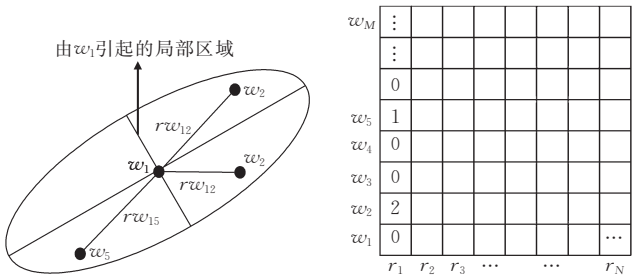


图 3　由特征单词 w_1 引起的区域关系及计数示意图

这样表示的好处有：

(1) 这种计数的表示方法可以反映区域的拓扑结构，因为使用的是“局部”特征单词. 也就是说这种关系是局部的，局部相对于全局来说是稳定的. 并且如果提取的特征单词是抗尺度和视角变化的，还可以对这些攻击具有很好的抵抗性. 实际应用中视角变换问题是一个难点，本文对区域的这种选择可以有效地抵抗这种干扰；

(2) 本文使用的特征单词的计数实际上反映了一种统计信息. 文献[1-4]表明，这种统计信息在物体分类识别中是一个很重要的因素. 因此，LLSR 方法将区域间的空间关系与特征单词的统计信息有机地结合在一起.

3.2　LLSR 与 part-based-methods 的区别

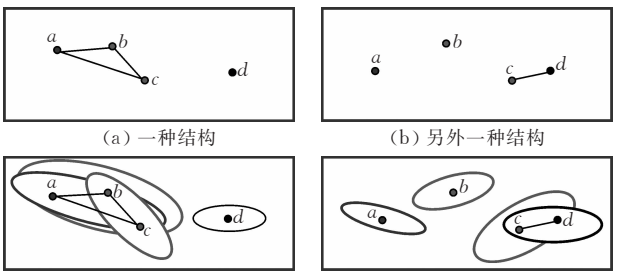
(1) 无监督与监督学习. Part-based-methods 往往需要明确地给出一个物体结构模型，并且需要人工参与. 这在实际大规模问题中是不可行的. 与之相反，LLSR 是一种无监督学习.

(2) 稠密模型与稀疏模型. Part-based-methods 是稀疏模型，这种模型很难抵抗遮挡问题. 对于几何攻击中的仿射变换也很难处理. 与之相反，LLSR 模型是稠密模型，由 3.1 节所述可知它可以有效抵抗几何变换、遮挡等难题.

(3) 本文的目的是要找到物体结构的一种潜在关系. 本文认为这种潜在的空间关系比明确的关系更能准确地描述物体. 这里涉及到人类认识学的问题，其中一个观点就是，被大家普遍认同的一种明确参数化的空间关系往往并不能真正地刻画这类物体，而某些潜在的非明确表达的空间关系能更准确地刻画物体.

3.3 LLSR 与 bag-of-word 方法的区别

两者之间最大差别是 LLSR 可以包含空间关系信息，而 bag-of-word 方法没有包含任何空间信息. 也就是说对于具有相同特征单词概率统计的不同图像 bag-of-word 是失效的，LLSR 方法可以对此加以区别. 例如，图 4(a)中的结构在(c)中以特征区域的方式表示，(b)中的结构在(d)中以特征区域的方式表示. 显然(a)和(b)具有不同的结构，但是具有相同的统计信息. bag-of-word 方法不能对其区分，而 LLSR 模型是可以区分的. 对于(a)和(b)，在 bag-of-word 中 a,b,c,d 点的统计信息都为 $\#a:\#b:\#c:\#d=1:1:1:1$. 然而在 LLSR 模型中，图 4(a)和(b)具有不同的结构，分别对应于(c)和(d). 相应的计数关系由(e)和(f)给出. 这种计数关系是截然不同的，LLSR 可以进行区分.



(a) 一种结构　(b) 另外一种结构

(c) 对应于(a)的区域关系　(d) 对应于(b)的区域关系

	r_a	r_b	r_c	r_d
w_d	0	0	0	1
w_c	1	1	1	0
w_b	1	1	1	0
w_a	1	1	0	0

(e) 与(c)对应的区域计数

	r_a	r_b	r_c	r_d
w_d	0	0	1	1
w_c	0	0	1	1
w_b	0	1	0	0
w_a	1	0	0	0

(f) 与(d)对应的区域计数

图 4　LLSR 对不同结构的区别

4 实　现

4.1 特征单词与特征区域的确定

我们采用多尺度抗仿射 Harris-Affine[14] 特征点方法提取图像边缘上的特征点. 只提取边缘上特征点的原因是为了消除歧义点影响. 该方法可以在不同尺度下提取出局部特征. 然后提取局部特征的描述符，本文使用 GLOH 方法[15]. 特征区域的窗口大小由特征尺度决定，通常以 3σ 为半径即可，其中 σ 为特征尺度. 每个特征区域需要旋转到其主方向上，然后映射为一个归一化的圆，使用 GLOH 提取特征描述. 这样就提供了一种抗平移、旋转、缩放、仿射变换的特征区域. 使用 K 平均方法对特征描述符进行聚类，形成特征码书. 这样每个特征点可以有相应的索引找到其对应的特征单词. 这里特征码书的大小是一个可以进行调解的参数. 码书太小则对区域的区分能力差，太大则会使相似的区域不能归并到一起. 实验中码书大小设为 500.

由前面叙述可知，每个特征区域是由特征单词引起的，因此每个特征单词会引起一个区域. 通过 Harris-Affine 方法可以确定特征的尺度，该尺度可以确定一个区域，通常为一个椭圆. 该区域内的若干特征单词与引起该区域的单词决定了该区域的结构. 具体可以参考本文 3.1 节.

4.2 学习与分类识别

学习主要包括从训练图像中建立区域-特征单词-图像三维表，然后使用 2.3 节中的 VEM 算法对参数进行估计. VEM 进行随机初始化，通常迭代 50～130 次即可收敛.

对于新的图像进行分类识别是一个贝叶斯决策过程. 首先需要对图像进行特征点提取，进而进行特征区域表示，根据特征码书划定特征单词. 设图像 d 具有 N 个区域，码书由 M 个特征单词组成，则需要判断后验概率 $P(z_k/d,\eta_1,\eta_2,\eta_3)$ 来决定当前图像的类别. 如果每种类别的先验概率均等，则有下式成立：

$$\begin{aligned}&P(z_k/d,\eta_1,\eta_2,\eta_3)\propto\\&P(d/z_k,\eta_1,\eta_2,\eta_3)\propto P(d/\eta_1,\beta,\theta,z_k)=\\&P(\gamma/\eta_1)\prod_{i=1}^{N}P(z_k/\gamma)P(r_i/\beta,z_k)\prod_{j=1}^{M}P(w_j/r_i,\theta)\end{aligned}\tag{10}$$

因此判断该图像的类别需要将 γ 进行边缘积分：

$$\begin{aligned}&P(d/z_k,\eta_1,\eta_2,\eta_3)\propto\\&\int P(\gamma/\eta_1)\prod_{i=1}^{N}P(z_k/\gamma)P(r_i/\beta,z_k)\prod_{j=1}^{M}P(w_j/r_i,\theta)\mathrm{d}\gamma\end{aligned}\tag{11}$$

得到最终判决公式，

$$\begin{aligned}k=\arg\max_k\int P(\gamma/\eta_1)\prod_{i=1}^{N}P(z_k/\gamma)P(r_i/\beta,z_k)\cdot\\\prod_{j=1}^{M}P(w_j/r_i,\theta)\mathrm{d}\gamma\end{aligned}\tag{12}$$

图 5 给出基于潜在局部区域空间关系模型的学习和分类过程.

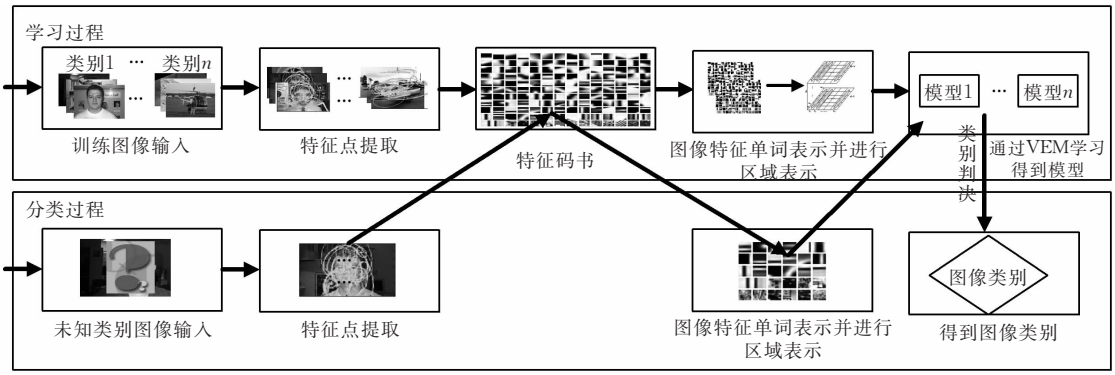


图 5　算法主要过程

5　实验对比与分析

5.1　图像数据

实验中的图像数据使用 Caltech 256 数据库①，该数据库是对 Caltech 101 的扩充，包含 257 组图像类别. 主要包括：435 张人脸图像；800 张摩托车图像；1150 张汽车图像；800 张飞机图像；900 张背景图像. 该数据库被各种算法广泛采用，因此具有较好的参照性，同时由于该图像库每种图像较多，实验过程比较方便.

图像大小也是一个考虑因素. 图像过小提取的特征点不足，会造成图像特征不能被真实表达，而图像过大由于提取的特征增多会造成很多冗余计算，表 1 给出了不同大小图像提取特征点数目的统计. 实验表明，图像归一化为 300×300 像素大小可以取得较好效果.

每一类图像都随机地分成两部分，一部分为训练图像；一部分为测试图像. 特征码书一旦训练好是固定的，我们通过训练图像集合得到特征码书.

表 1　不同大小图像提取特征点数目

图像大小	特征点数目
200×200	约 200
300×300	约 350
400×400	约 800

5.2　二类及多类结果

对于二类图像分类问题，模型中主题的数目为 2. 该问题也称为“物体-背景”问题. 实验中将包含主题的图像 200 幅，背景的图像 200 幅作为训练图像进行训练. 图 6 中给出了一个分类的例子，特征区域以椭圆方式标出.

在多类问题中，训练模型的主题数目为多个. 通过对训练图像的无监督学习，可得到每类图像的模型. 针对该问题，实验中进行了 5 类问题的实验，图像类别分别为人脸、汽车、摩托车、飞机、手表. 表 2 给出了实验结果. 我们通过混合表来评价算法，表中第 i 行第 j 列数值代表在第 i 类别分类问题中，将其分类为第 j 类的数目. 例如对表 2 第 1 行来讲，200 幅人脸图像中正确分类出 190 幅，在另外 10 幅中有 1 幅误分类为汽车，2 幅误分类为摩托车，2 幅误分类为飞机，5 幅误分类为手表. 表中对角线元素越大越好. 实验表明，本文算法可以有效地进行多类物体的分类.

表 2　多类分类问题结果的混合表

图像	人脸	汽车	摩托车	飞机	手表
人脸	190/200	1/200	2/200	2/200	5/200
汽车	0/200	193/200	5/200	1/200	1/200
摩托车	3/200	10/200	182/200	4/200	1/200
飞机	3/200	2/200	6/200	188/200	1/200
手表	7/200	4/200	8/200	2/200	179/200

5.3　具有相同统计信息的不同类图像的结果

如本文前面所述，LLSR 模型中不但具有统计信息，还具有刻画物体空间关系的性质. 因此具有相同统计信息的不同类图像使用本文算法可以有效进行分类. 与此相比较，bag-of-word[1-4] 方法会在这种情况下失效. 为了有效地进行验证，实验中将一半的待分类图像随机的进行分块重组，经过重组以后的图像已经不具有原图像的语义，因此这些图像应该归属为背景图像类.

实验中将 LLSR 方法与 Bayes[1] 和 pLSA[2] 两种经典方法进行了比较. 通过表 3 可以看出，三种算法对人脸图像都有较高的分类正确率. 由于没有包含空间信息，Bayes 和 pLSA 方法对背景的分类是失效的，而 LLSR 由于加入了空间信息能较好地对背景图像进行分类(对应于表 3 的第 3 行).

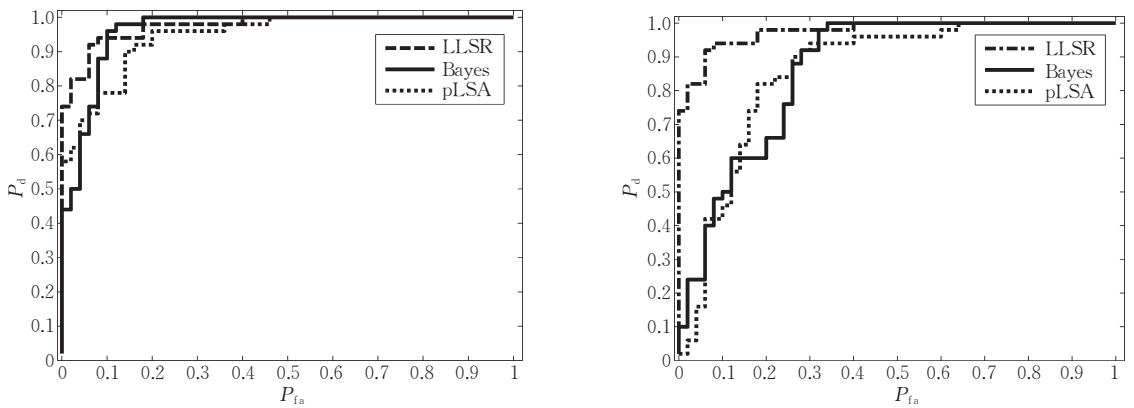
① http://www.vision.caltech.edu/Image_Datasets/Caltech256/

图 6 二类分类实验(每组给出 16 幅结果,并对分类结果进行了标注,正确和错误的分类分别以 correct 和 incorrect 标出)

表 3 空间信息对 LLSR、Bayes[1] 以及 pLSA[2] 三种方法的影响

图像	LLSR	Bayes	pLSA
正确分类的人脸	95/100	92/100	87/100
错误分类的人脸	5/100	8/100	13/100
正确分类的背景	**81/100**	**38/100**	**42/100**
错误分类的背景	19/100	62/100	58/100

此外,在正常图像的分类实验中,发现 LLSR 对图像大小不敏感,而 Bayes 和 pLSA 会随着图像大小的变化产生不规律的变化. 实验采用 ROC (Receiver-Operation Characteristic)曲线进行比较(ROC 曲线采用正确检测率 P_d 和错误检测率 P_{fa} 曲线表示,其中 P_d = 正确分类正样本数目/全部正样本数目,P_{fa} = 将负样本误认为正样本数目/全部负样本数目). 曲线下的面积越大算法越好. 图 7 是针对 200×200 和 400×400 两种情况下的 ROC 曲线.



(a) 图像大小200×200对应的ROC曲线 (b) 图像大小400×400对应的ROC曲线

图 7 LLSR、pLSA 和 Bayes 三种方法鲁棒性比较

如图 7 所示，对于第一种情况三种方法差别并不是很明显，但是随着图像的增大，提取的特征单词数目会增加，这对 LLSR 影响不大，相反 Bayes 和 pLSA 方法性能下降比较快，表明了 LLSR 具有更强的鲁棒性.

5.4　几何变换及遮挡条件下的结果

通过前述可知 LLSR 是抵抗几何变换的，由于模型是稠密的，因此在一定程度上对于遮挡问题也是有效的. 为此我们将 LLSR 方法与典型的几种 part-based-methods 进行了比较. 实验采用二类分类进行验证. 将待测图像进行平移、旋转、缩放、仿射和遮挡等操作. 这些条件中平移和旋转对各种方法结果影响不大. 因此，只给出缩放、仿射和遮挡条件下的结果. 比较标准为查全率(*recall*)r 与查准率(*precision*)p，它们定义为如下

$$r=\frac{\#\text{ 正确分类的个数}}{\#\text{ 此类别个数}} \tag{13}$$

$$p=\frac{\#\text{正确分类的个数}}{\#\text{被分为此类别的个数}} \tag{14}$$

相应的实验结果在表 4 中给出，可以看出，由于 LLSR 采用抗几何变换的局部特征点方法，因此对几何变换具有很好的结果. 即使仿射变换(affine)这种比较困难的情况，LLSR 也可以很好地处理. 与之比较的其它方法对仿射变换基本是失效的. 对于遮挡难题，稠密模型的优点也较 part-based-methods 有较好的提高.

综合前述各种方法，表 5 中给出了这些方法的一些特性，从中可以清楚地分析各种算法的优缺点.

表 4　不同的条件下 LLSR 与 Scale invariant[7] 和 Pictorial structures[9] 方法对二类分类查全率与查准率的实验结果

条件	方法	r/p				
		人脸	汽车	摩托车	飞机	手表
尺度变化	LLSR	0.885/0.908	0.770/0.758	0.926/0.958	0.826/0.880	0.930/0.930
	Scale invariant	0.766/0.763	0.580/0.513	0.907/0.924	0.622/0.610	0.886/0.930
	Pictorial structures	0.332/0.515	0.212/0.432	0.411/0.468	0.440/0.563	0.420/0.470
仿射变换	LLSR	0.885/0.908	0.792/0.776	0.930/0.945	0.742/0.810	0.890/0.902
	Scale invariant	0.225/0.271	0.283/0.422	0.351/0.557	0.292/0.557	0.612/0.583
	Pictorial structures	0.220/0.321	0.191/0.381	0.299/0.520	0.310/0.720	0.471/0.660
部分遮挡	LLSR	0.622/0.563	0.582/0.613	0.750/0.786	0.672/0.650	0.810/0.850
	Scale invariant	0.468/0.568	0.331/0.578	0.411/0.510	0.390/0.700	0.556/0.510
	Pictorial structures	0.560/0.522	0.412/0.452	0.398/0.499	0.422/0.469	0.600/0.710

表 5　不同算法特性的比较

特性	模型的表示	稠密/稀疏模型	是否具有统计框架	非监督/监督学习	全局/局部模型	能否抵抗平移	能否抵抗旋转	能否抵抗尺度变化	能否抵抗仿射变换	能否抵抗遮挡	空间信息	复杂度
LLSR	潜在	稠密	是	非监督	局部	能	能	能	能	部分	有	低
Bayes[1]	潜在	稠密	是	非监督	局部	能	能	能	能	困难	没有	低
pLSA[2]	潜在	稠密	是	非监督	局部	能	能	能	能	困难	没有	低
Scale invariant[7]	明确	稀疏	是	非监督	全局	能	能	能	不能	不能	有	中等
Pictorial structures[9]	明确	稀疏	是	监督	全局	能	能	部分	不能	不能	有	高
1-fans[10]	明确	稀疏	是	监督	全局	能	能	不能	不能	不能	有	低
2-fans[10]	明确	稀疏	是	监督	全局	能	能	能	不能	不能	有	中等
3-fans[10]	明确	稀疏	是	监督	全局	能	能	能	不能	不能	有	中等

6　结论与展望

将物体分类中的统计信息与空间信息相结合，本文提出了潜在局部区域空间关系模型，进而给出具体的实现算法. 模型参数的更新通过变分期望值最大(VEM)方法进行学习. 传统方法对空间信息的描述仅仅给出一个显示模型，这在实际问题的解决上受到了极大的限制，使得这些方法对很多情况不能很好地处理. 本文算法将这种显式的空间模型替换为一种通过局部特征点关系描述的潜在的空间关系. 本文算法将统计信息与空间信息很好地结合起来，在保留物体空间信息的同时对几何变换具有很好的抵抗性. 同时由于本文提出的模型是稠密模型，因此可以抵抗部分的遮挡问题. 与之比较的同类算法，往往侧重其中一方面. 与此同时，实验表明算法的计算复杂度并没有提高.

未来工作包括：(1)对参数的更新寻找更为有效、快速的学习方法；(2)同一图像中包含多个类别物体的问题需要进一步研究；(3)算法应用于视频数据的分类也是一个值得探究的问题.

参　考　文　献

[1]　Csurka G, Dance C, Fan L, Willamowski J, Bray C. Visual

categorization with bags of keypoints//Proceedings of the 2004 ECCV International Workshop on Statistical Learning in Computer Vision. Prague, Czech Republic, 2004: 59-74

[2] Sivic J, Russell B, Efros A, Zisserman A, Freeman W. Discovering objects and their localization in images//Proceedings of the 10th International Conference on Computer Vision (ICCV'05). Beijing, China, 2005, 1: 370-377

[3] Winn J, Criminisi A, Minka T. Object categorization by learned universal visual dictionary//Proceedings of the 10th International Conference on Computer Vision (ICCV'05). Beijing, China, 2005, 2: 1800-1807

[4] Li Fei-Fei, Fergus R, Perona P. One-shot learning of object categories. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 594-611

[5] Burl M, Weber M, Perona P. A probabilistic approach to object recognition using local photometry and global geometry//Proceedings of the 5th European Conference on Computer Vision. Freiburg, Germany, 1998, 2: 628-641

[6] Weber M, Welling M, Perona P. Unsupervised learning of models for recognition//Proceedings of the 6th European Conference on Computer Vision. Dublin, Ireland, 2000, 1: 18-32

[7] Fergus R, Perona P, Zisserman A. Object class recognition by unsupervised scale-invariant learning//Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03). Madison, Wisconsin, USA, 2003, 2: 264-271

[8] Agarwal S, Roth D. Learning a sparse representation for object detection//Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark, 2002, 4: 113-130

[9] Felzenszwalb P, Huttenlocher D. Pictorial structures for object recognition. International Journal of Computer Vision, 2005, 61(1): 55-79

[10] Crandall D, Felzenszwalb P, Huttenlocher D. Spatial priors for part-based recognition using statistical models//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, California, USA, 2005, 1: 10-17

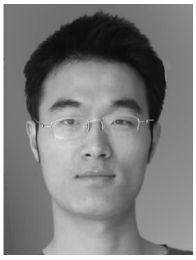
[11] Wainwright M, Jordan M. Graphical models, exponential families, and variational inference. Department of Statistics, University of California, Berkeley, California: Technical Report 649, 2003

[12] Blei D, Ng A, Jordan M. Latent Dirichlet allocation. Journal of Machine Learning Research, 2003, 3(3): 993-1022

[13] Frey B, Jojic N. A comparison of algorithms for inference and learning in probabilistic graphical models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(9): 1392-1416

[14] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. International Journal of Computer Vision, 2004, 60(1): 63-86

[15] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630



HAN Dong-Feng, born in 1979, Ph. D. candidate. His research interests include computer vision, machine learning and pattern recognition.

LI Wen-Hui, born in 1961, Ph. D., professor, Ph. D. supervisor. His research interests include computer vision and virtual reality.

GUO Wu, born in 1972, Ph. D. candidate. Her research interests include computer graphics and virtual reality.

Background

The problem of image classification is to categorize images using some image features and learning methods. The goal is to correctly classify unseen objects. Though this basic ability is easy and natural for humans, it is more difficult for computers to finish this work. A rich palette of diverse ideas has been proposed during the past few years. One of the most difficult problems is how to combine the statistics information with the spatial information to make the algorithm invariant to translation, rotation, scale, affine and occlusion.

In this paper, the authors focus their attentions on developing statistical methods which learn object spatial and appearance models from the training examples using unsupervised learning algorithm. LLSR model is proposed by analyzing the latent local spatial relations. The local count number is used to construct the relations of neighbourhood regions. This representation contains the relations of local regions and the statistics of visual words. LLSR is a hierarchical probabilistic model and the model parameters are learnt using variational expectation maximization. The variational methods are promising for large scale problems. Experiments show that the proposed method can capture both spatial relations and statistics information simultaneously and can resist to translation, rotation, scale, affine and part of occlusion.

This research is supported by the National Natural Science Foundation of China (grant Nos. 50338030, 60573182), the Doctor Foundation of China (grant No. 20060183042) and the Jilin Science Foundation (grant Nos. 20040531, 20060527). The research group has been working on computer vision areas for many years and has obtained some achievements in related fields. The group has published more than sixty papers in the international journals and conferences. More than twenty papers are indexed by SCI/SCIE and more than thirty papers are indexed by EI.