

基于内容相关性的场景图像分类方法

秦磊^{1,2} 高文^{2,3}

¹(中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)

²(中国科学院计算技术研究所 北京 100190)

³(北京大学数字媒体研究所 北京 100871)

(lqin@jdl.ac.cn)

Scene Image Categorization Based on Content Correlation

Qin Lei^{1,2} and Gao Wen^{2,3}

¹(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

³(Institute of Digital Media, Peking University, Beijing 100871)

Abstract Scene image categorization is a basic problem in the field of computer vision. A content correlation based scene image categorization method is proposed in this paper. First of all, dense local features are extracted from images. The local features are quantized to form visual words, and images are represented by the “bag-of-visual words” vector. Then a logistic-normal distribution-based generative model is used to learn themes in the training set, and themes distribution on each image in the training set. Finally, an SVM based discriminative model is used to train the multi-classifier. The proposed approach has the following advantages. Firstly, the approach uses logistic normal distribution as the prior distribution of themes. The correlation of themes is induced by the covariance matrix of logistic normal distribution, which makes the theme distribution of subjects more accurate. Secondly, manually tagging image content is not required in learning process, so as to avoid the heavy human labor and subjective uncertainty introduced in the process of labeling. A new local descriptor is proposed in this paper, which combines the gradient and color information of local area. Experimental results on natural scene dataset and manmade scene dataset show that the proposed scene image categorization method achieves better results than traditional methods.

Key words scene image categorization; generative model; logistic normal distribution; content correlation; computer vision

摘要 场景图像分类是计算机视觉领域中的一个基本问题。提出一种基于内容相关性的场景图像分类方法。首先从图像上提取视觉单词,并把图像表示成视觉单词的词频向量;然后利用产生式模型来学习训练集中包含的主题,和每一幅图像所包含的相关主题;最后用判定式分类器进行多类学习。提出的方法利用 logistic 正态分布对主题的相关性进行建模,使得学习得到的类别的主题分布更准确。并且在学习过程中不需要对图像内容进行人工标注,还提出了一种新的局部区域描述方法,它结合了局部区域

收稿日期:2007-10-22;修回日期:2009-02-11

基金项目:国家“八六三”高技术研究发展计划基金项目(2006AA01Z117);国家“九七三”重点基础研究发展计划基金项目(2009CB320900)

的梯度信息和彩色信息,在自然场景图像集合和人造场景图像集合上实验了提出的方法,它相对于传统方法取得了更好的结果。

关键词 场景图像分类;产生式模型;logistic 正态分布;内容相关性;计算机视觉

中图分类号 TP391.41

随着高分辨率数字照相机的普及,高质量数字图像的数量呈现出爆发式的增长。为了有效地管理和检索网络上及个人主机上的图像,对图像分类技术的需求越来越迫切。场景图像分类技术是其中重要的一类,其目标是把场景图像自动地分到有语义的场景类别中,如森林,海岸,公路等。图像依场景分类后可以有效地组织和浏览图像数据库,所以场景图像分类技术可以应用到如下领域中:个人影集或专业图像集合管理系统,基于内容的图像检索等。

根据是否需要首先识别场景中所出现的物体,可以把场景分类技术分为 2 大类。传统的观点认为,为了理解一个复杂的场景,必须先对场景中的物体进行识别,然后才能根据物体识别的结果理解场景,并识别场景的类别^[1]。但是如何有效地识别物体至今还没有解决,所以最近很多研究者尝试不进行物体识别,而是先构造场景的内容表示,然后根据场景的内容表示对场景分类^[2-3]。Oliva 等人提出了一个场景图像的低维表示方法,共有 5 维:自然性、开放性、粗糙性、辽阔性、险峻性^[2]。这个表示方法结合了场景的全局属性和局部属性对场景内容进行表述。Vogel 等人用一种通过人工标注得到的内容表示来进行场景分类^[3]。他们首先用一个均匀的网格把图像分成小区域,每个小区域被人工赋予一个概念,如岩石、草地等。然后图像就可以用一个内容概念矢量来表示。可以看到,文献[2-3]通过人工标注得到场景的内容表达,其中封装了内容概念间的相关性,但是它们有大量的数据需要进行人工标注:在文献[2]中需要给每个训练图像的 5 个不同属性进行人工打分;在文献[3]中需要把从训练图像中得到的 59582 个局部图像片人工分类到 9 个内容概念中。为了避免繁重的人工标注,最近有人尝试直接从图像中学习内容表达。Li 等人 and Bosch 等人分别利用 latent dirichlet allocation(LDA)^[4]模型和概率隐含语义分析模型(pLSA)^[5]进行无监督的场景分类^[6-7]。pLSA 和 LDA 最早被提出用来发现文本中所包含主题,它们都是以“bag-of-words”文本表示方法为基础的无监督分类算法。文献[6-7]的优点是它们利用了无监督分类算法,所以训练数据不需要人工标注,成功

地进行了无监督的场景分类。但是在它们的模型中图像内容之间是相互独立的,没能描述图像内容间的相关性,而事实上在场景图像中,图像内容是很强的相关性的。例如图 1(a)中包含 4 个区域,其中 3 个区域的分别是天空、山脉、大海,右下角是未知区域;图 1(b)是候选的未知区域,分别是沙滩、树林、田地、公路、厨房。则根据常识候选区域出现的概率是不同的,和天空、山脉、大海相关的沙滩、树林出现在图 1(b)中的概率较大,而和天空、山脉、大海不相关的公路、厨房出现在图 1(b)中的概率较小。

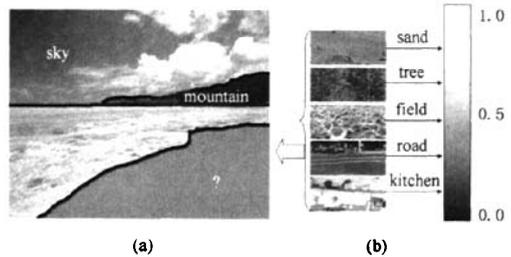


Fig. 1 The correlation of image content. (a) An image that contains four regions. The themes of three regions are sky, mountains and sea respectively. The theme of the lower right area is unknown and (b) The themes probability.

图 1 图像内容相关性示意图。(a) 一幅包含 4 个区域的图像,其中 3 个区域的主题分别是天空、山脉、大海,右下角是未知区域;(b) 未知区域的可能主题及其可能出现的概率

为了能同时拥有上述 2 种方法的优点,即在不需人工标注的情况下还能利用场景图像内容的相关性。在本文中,我们提出一种新的结合了产生式模型和判定式模型的弱监督场景分类方法。首先把图像表示成视觉单词的词频矢量,然后利用基于 logistic 正态分布的产生式模型把场景图像表示成主题分布矢量,并用支撑向量机分类器进行分类。在产生式模型中采用 logistic 正态分布作为场景图像主题的先验分布是因为通过 logistic 正态分布的协方差矩阵可以描述主题之间的相关性。

图 2 是我们方法的整体框架图。

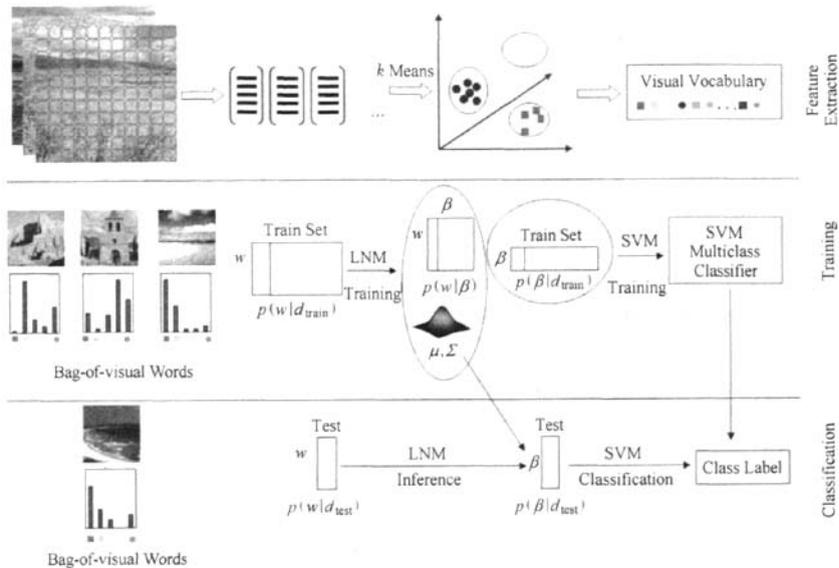


Fig. 2 The framework of the proposed scene image classification method.

图2 方法的框架图(由特征提取并构造视觉词汇表、训练和分类3部分组成)

1 视觉单词和视觉词汇表

最近的研究表明局部特征在很多计算机视觉领域取得了很可观的成功,例如图像匹配^[8]、物体识别^[9]、视频检索^[10]等领域。同全局特征相比,局部特征对遮挡和空间变化有很好的鲁棒性。根据在图像上的分布及提取方式,局部特征可以分为稀疏的局部特征和密集的局部特征。稀疏的局部特征由局部不变区域检测器来提取,主要分布在图像中特定结构附近,如 Harris-Affine 区域^[11]、DoG 区域^[12]等。密集的局部特征是对图像进行等间距的采样得到的局部区域,均匀分布在图像上。Li 等人在文献^[6]中的实验表明,密集的局部特征比稀疏的局部特征更适合于场景分类,所以下面我们用密集的局部特征来表示图像。

为了重复利用图像的色彩信息,我们提出一种新的局部特征表示方法——HSV 域上的梯度直方图(HSVGH),构造过程如下:

首先把图像转换到 HSV 空间上。对 HSV 空间的每个成分,在成分图像上叠加一个均匀的网格,网格的间距是 N 个像素。在每个网格点,从成分图像上提取半径为 R 的圆区域。在这个圆区域上计算 GLOH 特征^[13]。由于在场景分类任务中,多数情况下图像之间是没有明显的旋转变化的,所以我们去

掉了计算区域主方向的步骤。最后把 3 个成分上的 GLOH 特征连接,得到一个 384 维的特征。在这里我们取 $N=10, R=4, 8, 16$ 。 R 取 3 个值是为了保证对图像尺度变化的鲁棒性。

首先从训练集合上提取 HSVGH 特征集合,然后利用 k 均值算法对它们进行聚类。所有的聚类中心构成一个视觉词汇表 V , 每个聚类中心就是一个视觉单词。在聚类过程中,我们用的是欧氏距离。有了词汇表 V 后,对于每一个 HSVGH 特征,寻找离它最近的视觉单词 v_i , 并用视觉单词 v_i 表示这个局部特征。最后,每幅图像的“bag-of-visual words”表示就是它的视觉单词的统计直方图^[14]。

2 产生式模型

首先介绍要用到的术语和定义。

图像集和视觉词汇表:假定有 I 个图像 $D = \{d_1, d_2, \dots, d_I\}$, 其中每幅图像由 I_n 个图像块构成。把所有的图像块量化就形成了长为 m 的视觉词汇表 $W = \{w_1, w_2, \dots, w_m\}$ 。

场景类和主题:场景类指的是场景图像的分类如海岸、山脉等;主题指的是场景图像中所包含的如岩石、海水、天空等内容。一个主题 z 是一个在视觉词汇表上的多项式分布。

主题分布矢量:每幅图像 d 和一个主题分布矢量

$\{\theta_{d,i,k}\}$ 相连. 其中 $\sum_{i=1}^k \theta_{d,i} = 1, \theta_{d,i} \geq 0, \theta_{d,i}$ 反映了图像 d 中的视觉单词由第 i 主题产生的概率.

主题值: 一个视觉单词由一个主题产生. $z_{d,n}$ 表示图像 d 中的第 n 个视觉单词的主题.

图像的“bag-of-visual words”表示: 一幅图像中的每个图像块用其相应的视觉单词表示, 不考虑图像块在图像中的几何位置关系, 只统计相应视觉单词在图像中出现的次数. 这样一幅图像 d 就可以表示成一个长为 m 的矢量 $n_d = (n_{d,1}, n_{d,2}, \dots, n_{d,m})$, 其中 $n_{d,i}$ 表示第 i 个视觉单词在图像 d 中出现的次数. 这个矢量称为图像的“bag-of-visual words”表示.

2.1 基于 logistic 正态分布的产生式模型

为了能很好地建模场景图像内容的相关性, 我们利用 logistic 正态分布^[13] 作为图像主题分布矢量的先验分布. 它的协方差矩阵可以描述主题分布矢量成分间的相关性, 如图 3 所示, 分别为当主题成分之间无相关、负相关、正相关时的 logistic 正态分布. logistics 正态分布已经被用在文本语义挖掘领域中^[16-17].

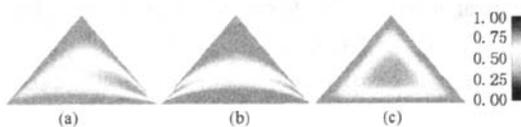


Fig. 3 The densities of logistic normal distribution on 2D simplex. (a) Diagonal covariance and nonzero-mean; (b) Negative correlation; and (c) Positive correlation.

图 3 在 2 维单纯形上的 logistic 正态分布密度. (a) 对角协方差矩阵; (b) 负相关; (c) 正相关

给定主题 $\beta_{1,k}, k$ 维矢量 $\mu, k \times k$ 协方差矩阵 Σ 后, 一个有 N 个视觉单词的图像 $d_i = \{w_{i,1}, w_{i,2}, \dots, w_{i,N}\}$ 的产生式过程如下.

1) 抽样图像 d_i 的主题分布矢量 θ_d, θ_d 是一个 k 维 logistic 正态随机变量.

① 抽样 $\alpha_d \sim N(\mu, \Sigma)$, 这里 $N(\mu, \Sigma)$ 是一个 k 维的正态分布;

② 通过下面的公式把 α_d 映射到单纯形上:

$$\theta_{d,k} = \exp\{\alpha_{d,k} - C(\bar{\alpha}_d)\};$$

$$C(\bar{\alpha}_d) = \ln\left(\sum_k \exp(\alpha_{d,k})\right).$$

2) 产生图像 d_i 中每个视觉单词 $w_{d,i}$:

① 抽样主题值 $z_{d,i} \sim \text{Multinomial}(\theta_d)$;

② 抽样视觉单词 $w_{d,i} | \{z_{d,i}, \beta_{1,k}\} \sim \text{Multinomial}(\beta_{z_{d,i}})$.

这个过程可以表示成如图 4 所示的概率图模型, 图中的矩形方框是一个缩略表示, 表示其中的节点要被复制多次, 矩形方框的左下角的数字是复制的次数.

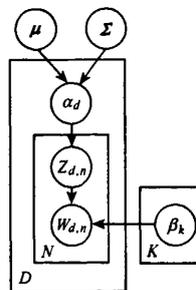


Fig. 4 Graphical model representation of the logistic normal based generative model.

图 4 基于 logistic 正态分布的产生式模型的概率图表示

2.2 模型的计算

给定一个图像 $d = \{w_1, w_2, \dots, w_N\}$ 和产生式模型参数 $\{\beta_{1,k}, \mu, \Sigma\}$, 则图像隐含变量的后验概率是:

$$p(\alpha, z | d, \beta_{1,k}, \mu, \Sigma) = \frac{p(\alpha | \mu, \Sigma) \prod_{n=1}^N p(z_n | \alpha) p(w_n | z_n, \beta_{1,k})}{\int p(\alpha | \mu, \Sigma) \prod_{n=1}^N \sum_{z_n=1}^K p(z_n | \alpha) p(w_n | z_n, \beta_{1,k}) d\alpha}$$

因为分母上的积分所以不能直接计算, 但是可以利用很多数值计算方法来近似计算这个后验概率, 如 MCMC 方法、Laplace 近似和变分方法^[18]. 在本文中, 我们利用变分方法来进行近似计算.

我们用 Jensen 不等式来界定一个图像的对数概率:

$$\log p(d | \mu, \Sigma, \beta) \geq E_q(\ln p(\alpha | \mu, \Sigma)) + \sum_{n=1}^N E_q(\ln p(z_n | \alpha)) + \sum_{n=1}^N E_q(\ln p(w_n | z_n, \beta)) + H(q), \quad (1)$$

这里 $q(\alpha, z | \lambda, v^2, \phi)$ 是一个变分分布, $H(q)$ 是这个分布的熵. 用 $J(q)$ 表示式 (1) 的右侧, 则: $\log p(d | \mu, \Sigma, \beta) = J(q) + KL(q(\alpha, z | \lambda, v^2, \phi) \| p(\alpha, z | d, \mu, \Sigma, \beta))$. 所以我们最大化下界 $J(q)$ 就相当于最小化变分后验分布和真正后验分布的 Kullback-Leibler 距离.

给定式 (1) 后, 我们通过变分 EM 来对基于 logistic 正态分布的产生式模型进行参数估计^[16,18]. E 步 M 步重复直到似然函数的界收敛.

1) 在 E 步中, 通过对每个图像进行变分推导,

最大化相对于变分参数 λ, v^2, ϕ 的界。

2) 在 M 步中, 我们相对于模型参数 μ, Σ, β 最大化这个界, 迭代公式如下:

$$\beta_i \propto \sum_d \phi_{d,i} n_d,$$

$$\mu = \frac{1}{D} \sum_d \lambda_d,$$

$$\Sigma = \frac{1}{D} \sum_d I v_d^2 + (\lambda_d - \mu) (\lambda_d - \mu)^T,$$

这里 n_d 是图像 d 的词频矢量。

3 混合分类

训练分成 2 个阶段。首先, 用整个训练图像集合来训练基于 logistic 正态分布的产生式模型, 学习到 $\beta_{i,k}$ 和 $\theta_{1,d,train}$ 。这样每幅训练图像 j 可以用一个 k 维矢量 θ_j 来表示。在这个阶段不需要提供训练图像类别标签。在第 2 个阶段给定了每幅图像的矢量 θ 和图像的类别标签后, 用一个判别式的多类分类器来学习多个类别。我们选用的多类分类器是支持向量机(SVM)^[19]。SVM 的多类分类用的是对一的策略。在训练时, 假设我们有 K 个类, 就要构造 $K(K-1)/2$ 个二类分类器, 每个二类分类器用 2 个类的训练集合进行训练。在分类时, 对一幅测试图像用第 i 个二类分类器进行分类($i=1, 2, \dots, K(K-1)/2$), 赢的类别加 1 票。最后得票最高的类别就是这幅测试图像所属类别。图 5 是利用对一策略的 SVM 多类分类器示意图。

测试图像的分类过程也是分成 2 个阶段。首先, 计算测试图像特定的混合系数 θ_{cent} , 然后用 SVM 分

类器来对混合系数 θ_{cent} 进行分类。图 2 是我们方法的框架图。

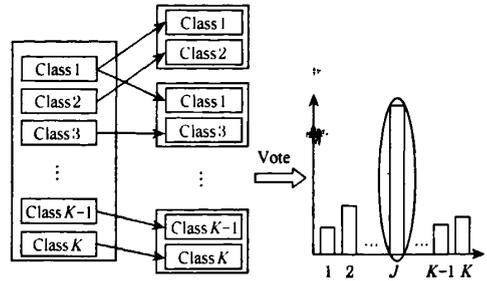


Fig. 5 SVM multi-class classifier which uses the one vs. one strategy.

图 5 利用对一策略构建 SVM 多类分类器

4 数据集及评价方法

在实验中, 我们利用了如下 2 个数据集。

D1: 由 Oliva 和 Torralba 提供的包含 4 类自然图像的数据集。数据集中共有 1472 幅彩色图像, 分别是野外 410 幅、海岸 360 幅、森林 328 幅、山脉 374 幅。图像的平均尺寸是 250×250 像素。图像的主要来源是 COREL 图像和个人影集。例子图像如图 6(a) 所示。

D2: 由 Oliva 和 Torralba 提供的包含 4 类人造图像的数据集。数据集中共有 1216 幅彩色图像, 分别是公路 260 幅、城市 308 幅、街道 292 幅、高楼 356 幅。图像的平均尺寸是 250×250 像素。例子图像如图 6(b) 所示。



Fig. 6 Example images of two datasets. (a) D1 consists of four nature scenes and (b) D2 consists of four manmade scenes.

图 6 2 个图像数据集的例子图像。(a) 4 类自然景色图像集合组成训练集 D1; (b) 4 类人造物体图像集合组成训练集 D2

在下一节, 我们给出在数据集 D1 和 D2 上的试验结果。从每类图像中随机选择 150 幅图像作为训练集合, 剩下的图像作为测试集合。分类任务是分给每幅测试图像一个类别值。分类完成后记录下每类的平均识别率。我们用混淆矩阵和全局分类率来衡量算法的性能。

5 实验结果

5.1 参数选择

在我们的方法中有 2 个自由参数: 1) 视觉词汇表的长度; 2) 主题的个数。为了调查这 2 个参数对

分类效果的影响,我们在数据集 D1 上进行了测试.测试的结果如图 7 所示.可以看到在视觉词汇表

的长度为 500,主题个数为 20 时,分类效果最好.在后面的实验中我们就用这 2 个数值.

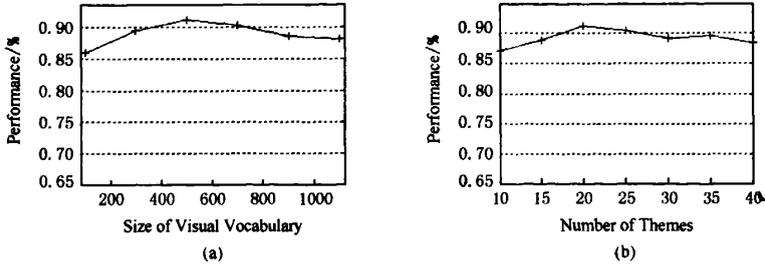


Fig. 7 The performance is influenced by the length of vocabulary and themes. (a) Performance vs. vocabulary and (b) Performance vs. themes.

图 7 改变视觉词汇表长度和隐含主题个数对数据集 D1 分类效果的影响. (a) 改变视觉词汇表长度时的分类性能变化曲线; (b) 改变隐含主题个数时的分类性能变化曲线

5.2 结果

我们报告了在数据集 D1 和 D2 上的实验结果.图 8(a)是数据集 D1 上的主题分布,图 8(b)是

数据集 D1 上的分类混淆矩阵.数据集 D1 上的平均分类率是 91.4%.图 9(a)是数据集 D2 上的主题分布,图 9(b)是数据集 D2 上的分类混淆矩阵.数据集

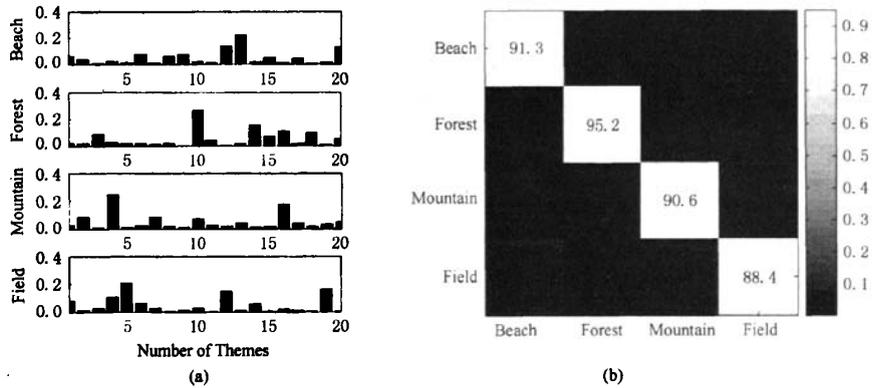


Fig. 8 Experiment results on D1. (a) Themes distribution of D1 and (b) Confusion matrix.

图 8 数据集 D1 上的实验结果. (a) D1 上的主题分布图; (b) D1 的分类混淆矩阵

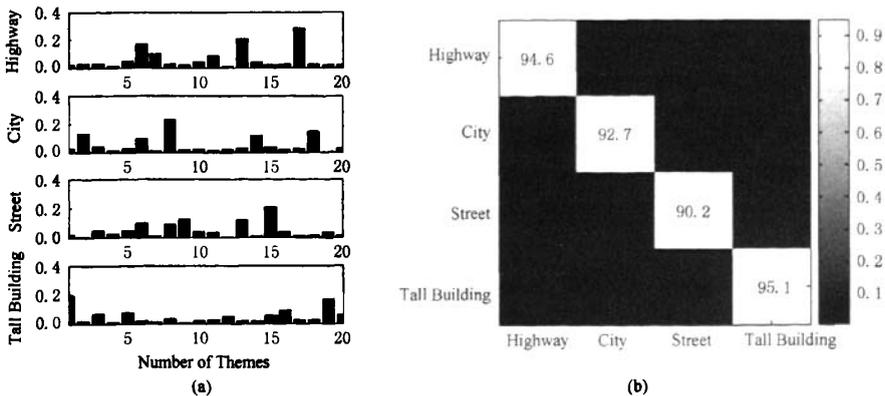


Fig. 9 Experimental results on D2. (a) Themes distribution of D2 and (b) Confusion matrix.

图 9 数据集 D2 上的实验结果. (a) D2 上的主题分布图; (b) D2 上的分类混淆矩阵

D2 上的平均分类率是 93.2%。

从图 8(a)和图 9(a)上我们可以看出,主题分布的能量是比较集中的.这是由于在我们的方法中引入了主题之间的相关性,加大了相关主题产生的概率,并且减小了不相关主题产生的概率.例如在公路类中,能量主要集中在 6,7,11,13,17 维上.在 D1 中分类结果最好的是森林,分类率是 95.2%,在 D2 中分类结果最好的是高楼,分类率是 95.1%.混淆比较大的类别是海岸和野外、山脉和野外、城市和街道。

图 10 给出了一些分类错误的场景图像.图 10(a)是野外图像被分为海岸图像,图 10(b)是街道图像

被分为城市图像.这些结果显示了有些场景类别之间容易混淆的原因是由于场景图像分类结果具有不确定性.而这种不确定性即使是在人工分类过程中也是不可避免的,所以增加了场景分类的难度.如图 10(a)中的野外图像同很多海岸图像一样有大面积的水面及远处的天空,所以完全可以把它们当成是海岸场景图像.图 10(b)中的街道图像也可以认为是城市图像,因为它们含有同样的城市建筑.还有 2 类容易混淆的类别是公路图像和街道图像,原因主要是街道图像一般都包含一条路,而这正是公路图像的主要内容。

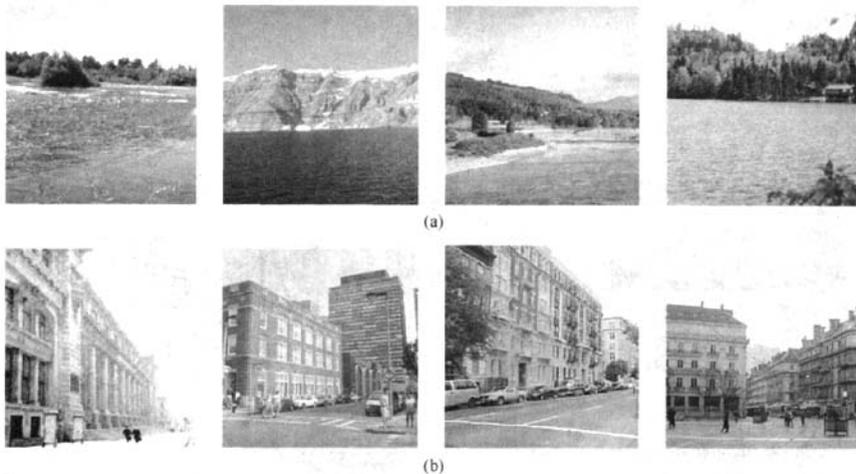


Fig. 10 Some misclassified images. (a) The images belong to field, which are misclassified as beach and (b) The images belong to street, which are misclassified as city.

图 10 一些分类错误的场景图像。(a) 野外被分为海岸;(b) 街道被分为城市

5.3 实验结果比较

本节比较了我们的方法和文献[2]中的方法.在文献[2]中,每类需要的训练图像个数是 250 幅到 300 幅,训练阶段需要人工标注训练图像的 5 个属性.在我们的方法中每类需要的训练图像个数是 150 幅,训练阶段不需要人工标注.从表 1 中,可以看到无论在 D1 还是 D2 上我们的方法都取得了更高的分类精度.我们的方法能取得更好效果的原因在于:1)我们的方法能描述场景内容的相关性;2)我们利用的特征有更好的表示性。

Table 1 Results Comparison with Related Work

表 1 我们的方法同其他方法分类结果的比较 %

Dataset	Categories	The Proposed Method	The Method of Ref [2]
D1	4 Nature Scene	91.4	89.1
D2	4 Manmade Scene	93.2	89.2

6 结 论

本文分析了当前基于内容的场景分类算法研究现状,针对存在的问题,提出一种基于内容相关性的场景图像分类方法.它有如下特点:

- 1)用 logistic 正态分布对主题的相关性进行建模.从而在学习过程中减少不相关主题出现的可能,并使得学习得到的类别的主题分布更准确.
- 2)在学习过程中不需要对图像内容进行人工标注,从而避免了繁重的人力劳动和标注过程中引入的主观不确定性.

在本文中,我们还提出了一种新的局部区域描述方法,它结合了局部区域的梯度信息和彩色信息.从实验结果来看,我们的方法成功地完成了场景图像分类任务,而且相对于传统方法取得了更好的结果。

参 考 文 献

- [1] Treisman A, Gelade G. A feature integration theory of attention [J]. *Cognitive Psychology*, 1980, 12(1): 97-136
- [2] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope [J]. *International Journal on Computer Vision*, 2001, 42(3): 145-175
- [3] Vogel J, Schiele B. Natural scene retrieval based on a semantic modeling step [C] //Proc of the ACM Int Conf on Image and Video Retrieval. New York: ACM, 2004: 207-215
- [4] Blei D, Ng A, Jordan M. Latent Dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(5): 993-1022
- [5] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis [J]. *Machine Learning*, 2001, 42(1): 177-196
- [6] Li Feifei, Perona P. A Bayesian hierarchical model for learning natural scene categories [C] //Proc of the IEEE Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE, 2005: 524-531
- [7] Anna Bosch, Andrew Zisserman, Xavier Munoz. Scene classification via pLSA [C] //Proc of the European Conf on Computer Vision, Berlin: Springer, 2006: 517-530
- [8] Qin Lei, Zeng Wei, Gao Wen, et al. Local invariant descriptor for image matching [C] //Proc of the IEEE Int Conf on Acoustics, Speech, and Signal Processing, Piscataway, NJ: IEEE, 2005: 1025-1028
- [9] Sivic J, Russell B C, Efros A A, et al. Discovering objects and their localization in images [C] //Proc of the Int Conf on Computer Vision, Washington, DC: IEEE Computer Society, 2005: 370-377
- [10] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos [C] //Proc of the Int Conf on Computer Vision, Washington, DC: IEEE Computer Society, 2003: 1470-1477
- [11] Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. *International Journal on Computer Vision*, 2004, 60(1): 63-86
- [12] Lowe D. Distinctive image features from scale-invariant keypoints [J]. *International Journal on Computer Vision*, 2004, 60(2): 91-110
- [13] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2005, 27(10): 1615-1630
- [14] Qin Lei, Zheng Qingfang, et al. Unsupervised texture classification: Automatically discover and classify texture patterns [J]. *Image and Vision Computing*, 2008, 26(5): 647-656
- [15] Aitchison J, Shen S M. Logistic-normal distributions: Some properties and uses [J]. *Biometrika*, 1980, 67(2): 261-272
- [16] Blei D, Lafferty J D. Correlated topic models [C] //Proc of Neural Information Processing Systems, Cambridge: MIT Press, 2006: 147-154
- [17] Ahmed A, Xing E P. On tight approximate inference of logistic normal admixture model [C/OL] //Proc of the Int Conf on Artificial Intelligence and Statistics, 2007 [2007-10-20]. <http://www.stat.umu.edu/~aistat/proceedings/data/papers/003.pdf>
- [18] Winn J. Variational message passing and its applications [D]. England: University of Cambridge, 2003
- [19] Bian Zhaoqi, Zhang Xuegong, et al. Pattern Recognition [M], 2nd ed. Beijing: Tsinghua University Press, 2000 (in Chinese)
(边肇祺, 张学工, 等. 模式识别[M], 2版. 北京: 清华大学出版社, 2000)



Qin Lei, born in 1977. Received his PhD degree in computer application from the Institute of Computing Technology, Chinese Academy of Sciences, in 2008. His main research interests include computer vision, pattern recognition and image processing.

秦磊, 1977年生, 博士, 主要研究方向为计算机视觉、模式识别和图像处理。



Gao Wen, born in 1956. Received his PhD degrees in computer science from Harbin Institute of Technology, China, in 1988, and in electronics engineering from the University of Tokyo, in 1991 respectively. Professor and PhD supervisor. Senior

member of China Computer Federation. His main research interests include multimedia data compression, image processing, computer vision, multimodal interface, artificial intelligence and virtual reality.

高文, 1956年生, 博士, 教授, 博士生导师, 中国计算机学会高级会员, 主要研究方向为多媒体数据压缩、图像处理、计算机视觉、多模式接口、人工智能和虚拟现实等。

Research Background

Scene categorization is a fundamental problem in computer vision. Automatic techniques for tagging scene images with semantic labels have a high potential for improving the performance of other computer vision applications such as retrieval, browsing and object recognition. The scene categorization problem has two critical components: representing scenes and training semantic categories models using these representations. In recent years, extensive research efforts have been devoted to scene categorization. Our work is partially supported by the National 863 High Technology Development Program of China under grant No. 2006AA01Z17, the National 973 Basic Research Program of China under grant No. 2009CB320900.