

基于 ML-pLSA 模型的目标识别算法

陈琳 卢湖川*

(大连理工大学信息与通信工程学院 大连 116024)

摘要: 为了避免图像目标识别过程中识别结果对分割结果的过度依赖, 该文提出了一种基于多尺度的概率潜在语义分析目标识别方法(Multi-Level-probabilistic Latent Semantic Analysis, ML-pLSA)。该方法利用多种分割算法对图像进行多尺度分割, 再利用 pLSA 算法和词袋方法(Bag Of Words, BOW)对分割区域进行目标类别估计, 最后联合多尺度的估计值给出最终分割结果。在目标尺度、目标角度、外界光照变化都相对较大的 GRAZ-02 数据库上进行实验, 结果表明: 与传统目标识别算法相比, 该方法鲁棒性更强; 在识别准确率方面, 也有了很大的提高, 取得了很好的效果。

关键词: 目标识别; 多种分割; 多尺度; 多尺度概率潜在语义分析(ML-pLSA)

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2011)12-2909-07

DOI: 10.3724/SP.J.1146.2011.00455

A New Object Recognition Method Based on ML-pLSA Model

Chen Lin Lu Hu-chuan

(School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: In order to avoid the condition that most of the segmentation based recognition methods have relied too much on the quality of image segments, a new object recognition method is proposed based on Multi-Level-probabilistic Latent Semantic Analysis (ML-pLSA) object recognition algorithm. Firstly, multiple segmentations at different levels are computed for each image, and then object classes on each segment region are estimated by using pLSA and Bag-Of-Words (BOW). The final results are obtained by fusing estimation results at multiple levels. The proposed algorithm is evaluated on Graz-02 dataset, a challenging dataset that contains large changes in object scale, object viewpoint and illumination condition. The experiment results demonstrate that the proposed method performs better than traditional object recognition methods in both accuracy and robustness.

Key words: Object recognition; Multi-segmentation; Multi-level; Multi-Level-probabilistic Latent Semantic Analysis (ML-pLSA)

1 引言

在图像分类和图像检索技术应用越来越广泛的今天, 图像目标识别技术成为了最热门的研究方向之一。图像大小、目标尺度、场景内容及光照、角度等因素的不确定性, 对目标识别算法构成了很大的挑战。现有的目标识别方法尽管取得了一定的成功, 但是也存在一定的问题: (1)由于目标尺度和形状的不确定性, 基于滑动窗的方法会引入大量的背景噪声, 且计算量太大。(2)在一系列先分割再识别的方法框架下, 当分割结果很不理想时, 后续的工作(无论边界/形状匹配)都是很难进行的。

针对以上问题, Russell 等人^[1]将多次改变计算参数的多分割方法用在了图像识别中, 其优点是多尺度分割可以避免目标尺度、图像大小的变化问题; 区域可以包含更丰富的特征信息。但是该方法的假设往往不成立, 因为即使分割的次数再多, 也有无法得到正确目标的情况发生。文献[2]提出了分割-识别-再分割的一个识别过程, 该算法避免了因为某一步分割结果的错误而导致整个识别结果失败的现象。但是该方法对格外突出的物体部分比较敏感(比如天鹅的脖子, 苹果商标的叶子等), 当突出部分形变较大时匹配困难且此方法对一般物体推广性还不够强。通过以上分析, 要做好目标识别必须解决以下几个问题: (1)尽量避免图像大小、目标尺度变化对识别的影响; (2)在尽量多的特征空间上描述目标; (3)能够稳定地对图像库的类别结构进行建模;

2011-05-16 收到, 2011-08-29 改回

国家自然科学基金(61071209)和中央高校基础研究基金(DUT10JS05)资助课题

*通信作者: 卢湖川 lhchuan@dlut.edu.cn

(4)识别过程与分割过程既要相互依赖又要相对独立。

鉴于此,本文提出一种基于多尺度的概率潜在语义分析(ML-pLSA)模型的目标识别算法。该算法选取多种分割方法对图像进行多尺度分割,然后利用词袋方法(BOW)结合 pLSA 模型及分类器对每一个分割区域进行类别估计,最终将多个尺度、多个方法的区域估计结果结合到一起,根据这一综合结果提取出目标,达到目标识别的目的。本文方法不需要假设分割出的区域至少有一个是正确的,也允许分类器有判断的错误。我们的每一步都是弱分类、弱学习,但是数据的叠加和相互依赖可以不断增强目标的位置信息和类别信息。

2 概率潜在语义分析(pLSA)

概率潜在语义分析(pLSA)模型^[3]原本是用文本检索的概率生成模型。相比标准潜在语义分析(LSA),pLSA 模型来自线性代数和执行奇异值分解共生表,是基于一个(从潜在的类模型的)混合体的分解,具有更牢固的数学基础。如图 1 所示。

图 1(a)中,节点被包含在的某个框中,表示该节点被迭代左上角的符号所标识的次数。实心圆表示观测到的随机变量;空心圆表示未观测到的随机变量。pLSA 的目的是找到特定主题中字的分布

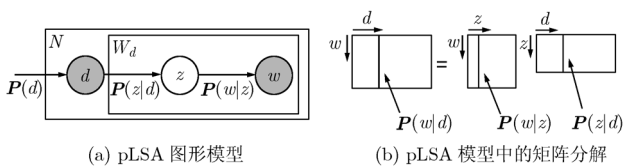


图1 pLSA 图形化模型

$P(w|z)$, 以及使特定文件中字的分布 $P(w|d)$ 组合起来的相应的特定文件的混合比例 $P(z|d)$, 如图 1(b)。pLSA 模型已在检索和信息过滤、自然语言处理、机器学习的文本和相关领域广泛应用^[4,5]。

3 基于 ML-pLSA 模型的目标识别算法

基于 ML-pLSA 模型的目标识别算法示意图如图 2 所示。其中,图 2(a)表示在不同的特征空间原图像突出不同的特征。图 2(b)用多种分割方法对图像进行分割,可以尽量多地利用不同特征空间信息。图 2(c)利用 pLSA 模型和分类器得到每个区域的置信值。图 2(d)表示多种分割方法的置信图的合并图。图 2(e)根据合并图提取的目标。从方法示意图可以看出,图 2(b)-(c)在简单分割的基础上进行识别判断。图 2(d)-(e)在一定识别的基础上进行提取、分割。整个识别过程中分割与识别相互依赖,位置与类别的信息在不断的增强。

3.1 特征提取

为了建立图像局部区域的描述,使其对视角、光照等的变化具有一定的鲁棒性,本文选择快速 SIFT(quick Scale Invariant Feature Transform, SIFT)描述符^[6],结合 BOW 进行直方图视觉词描述。本文以局部区域为样本,对区域内的像素点密集提取 SIFT 特征,然后形成区域特征直方图,这样可以更完整地描述局部特征。

3.2 多种分割算法的选择及改进

本文给每个输入图像产生足够的分割,为了能够产生尽量多样的好的区域(所谓的好区域就是指尽量多的包含目标)。但是有的图像偏重于颜色特征,而有些图像则更偏重的是纹理特征,若只用一

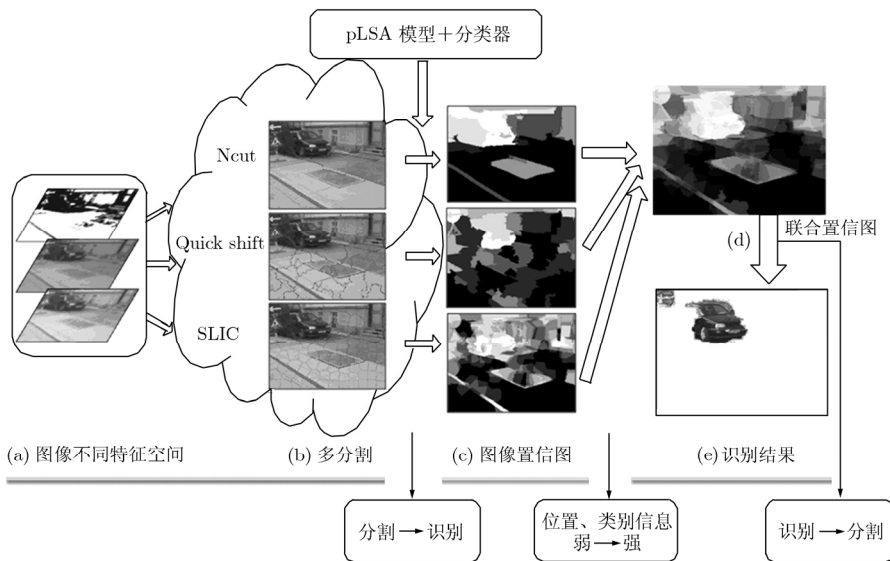


图2 ML-pLSA 方法示意图

种分割方法对图像库进行分割的话, 得到的分割效果参差不齐。所以本文决定选择多种分割方法对图像库的每张图像进行分割, 利用每种方法依赖的特征(cues)不同来弥补这方面的不足; 还选择在不同尺度上进行分割, 这样会防止由于目标大小不同而对分割产生的影响。这样一来, 得到的分割区域大小形状各有不同、依据的特征各有不同, 可以很好地克服因为不同图像库目标尺度变化大、图像特征复杂而带来的问题。

文献[7]中给出了目前较流行的分割方法的比较。综合数据, 本文选择归一化分割(Normalized cuts, Ncut)^[8]、快速漂移算法(Quick shift)^[9]、简单线性迭代聚类算法(SLIC)^[7] 3 种分割方法。Ncut 方法是基于全局最优的分割算法, 已被成功用于人体模型估计等领域^[10,11]。它对一次性分割出整个目标的可能性是最大的, 所以选择 Ncut 方法进行大尺度分割; SLIC 算法是 3 种方法中速度最快的, 而且由于他产生的超像素大小、形状基本相同, 不会过分注意一些没意义的拐角/线, 所以选择 SLIC 方法进行小尺度分割; 至于中间尺度的分割则由快速漂移算法来完成, 这样既可以发挥它的特点(把不规则的、有意义的区域分割出来), 又不会产生过小的超像素。快速漂移算法已广泛应用于图像识别和视频识别^[12,13]。根据他们的特点, 本文设定超像素的个数分别为 $K_{\text{ncut}}=[5,9,13,17,21,25]$, $K_{\text{quick}}=[43,41,39,37,35,33]$ (分割块数大概是 30-60), $K_{\text{slic}}=[70,80,90,100,110,120]$, 使分割块数基本覆盖了每张图片 5-120 块左右。

在快速漂移算法中, 本文将用于聚类的滤波器组^[14]用于提取局部纹理特征。滤波器组由 3 个高斯滤波器(方差 σ 分别为 1,2,4), 4 个拉普拉斯高斯滤波器(方差 σ 分别为 1,2,4,8), 以及 4 个高斯一阶导数滤波器。首先, 对输入图像进行颜色空间转换; 将输入图像由 RGB 颜色空间转换到 CIE-LAB 颜色空间; (1)用上述 3 个不同尺度的高斯核对 L, A, B 3 个颜色通道分别进行卷积产生 9 组滤波响应; (2)用 4 个不同尺度的拉普拉斯高斯滤波器仅仅对 L 通道进行卷积产生 4 组滤波响应; (3) 4 个高斯一阶导数滤波器对 L 通道进行滤波产生 4 组滤波响应; 所以最终每幅图像的每一个像素会得到一个 17 维的特征向量。滤波器的形状如图 3 所示。

3.3 ML-pLSA 模型

pLSA 模型原本是由于文本检索的概率生成模型。通过利用视觉词(量化的颜色特征, 纹理特征以及 SIFT 特征等区域描述子), pLSA 模型可以被应用于图像领域。本文尝试利用 pLSA 模型对分层分割区域进行分析并发现其中的“主题”, 把对象类别

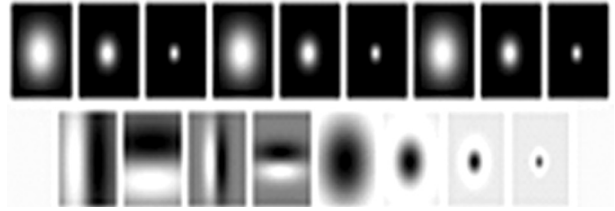


图 3 17 维滤波器组形状示意图

作为发现的“主题”(如草, 屋), 把图像包含多个对象实例建模为主题的混合物。对于每张图片可能存在多目标类的这种情况, pLSA 提供了正确用于聚类的统计模型。

下面, 运用图像处理的语言对 ML-pLSA 模型进行一下描述。ML-pLSA 模型中的原始术语“文件- d ”对应区域样本, “字- w ”对应区域样本特征, “主题- z ”对应目标类别, 它是一个潜在的中间变量。

假设有一组 M 个训练样本 $\{d_1, \dots, d_M\}$, 每个样本对应一个局部区域, 这些区域被量化为包含 W 个视觉单词的词汇表 $\{w_1, \dots, w_W\}$, 因此训练图像的集合就可以由一个单词图像的互共现矩阵 $N_{ij} = n(w_i, d_j)$ 来表示, 其中 $n(w_i, d_j)$ 表示的是文件 d_j 中字 w_i 出现的次数。假设共有 K 个潜在主题变量 $\{z_1, \dots, z_K\}$, 那么每个文件 d_j 中的每个字 w_i 的出现都有一个潜在的主题变量 z_k 与之相关联。

我们假设联合概率 $P(w_i, d_j, z_k)$ 拥有图 1(a)所示的图模型的形式。对主题 z_k 进行边缘求和确定出条件概率 $P(w_i | d_j)$:

$$P(w_i | d_j) = \sum_{k=1}^K P(z_k | d_j) P(w_i | z_k) \quad (1)$$

其中 $P(z_k | d_j)$ 为主题 z_k 在文件 d_j 中出现的概率; $P(w_i | z_k)$ 为字 w_i 在特定主题 z_k 中出现的概率。

式(1)将每个文件表示为 K 个主题向量的凸合并。这相当于进行一次图 1(b)所示的矩阵分解, 其中要求对向量和混合系数进行归一化从而使他们依概率分布。本质上说, 每个文件 d 都是不同主题 z 的混合物, 某个特定文件 d 的直方图是由每个主题 z 所对应的直方图相混合而组成的。

在具体实现过程中, 如图 4, 本文首先对 $P(w | d_{\text{train}})$ 进行奇异值分解, 得到降维后矩阵 $P(z | d_{\text{train}})$

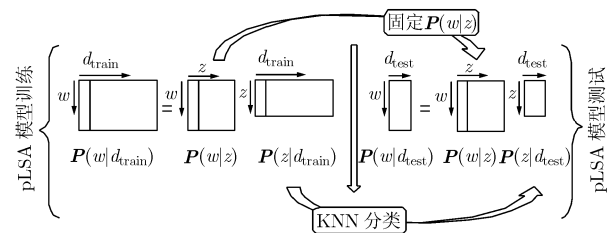


图 4 ML-pLSA 算法模型

和概率 $P(w|z)$ 。然后采用迭代操作, 固定概率 $P(w|z)$, 计算测试图像的潜在主题表示 $P(z|d_{\text{test}})$ 。最后对 $P(z|d_{\text{test}})$ 与 $P(z|d_{\text{train}})$ 进行相似性的度量。

4 实验结果

本文在常用的 GRAZ-02 数据库进行了一系列的实验。GRAZ-02 数据库包含 3 个类别: 车类共

300 张图片; 人类共 311 张图片; 自行车类共 365 张图片。每张图片大小 480×640 。示例图片如图 5 所示。可见, GRAZ-02 图像库中包括大目标、小目标、目标遮挡、多目标、光照变化、角度旋转(分别对应图 5 中第 1 列至第 6 列)等一系列图片, 对目标识别具有很大的挑战。



图 5 GRAZ-02 数据库实例图

我们选择图像库每类单数的 150 张图片来生成字典, 同时总共取出 100 张图片作为验证集合。在验证集合上我们得到的最优字典大小是 70 万至 80 万左右。本文每张图片提取 5000 个样本点(正负各半)的 SIFT 特征, 然后对矩阵进行 k -means 聚类, $K=400$ 。

选择每类图像的单数 150 张图片进行训练, 双数的所有图片进行测试。对于训练图像, 先以一种分割的一种参数(以分割块数 $N_{\text{cut_p}}=5$ 为例)对图片进行分割, 然后在每一个区域中提取 SIFT 特征, 将每一个区域的 SIFT 特征直方图投影到字典上, 得到 $K \times 5$ 维的区域特征直方图。通过计算每个区域中正/负像素点的个数给出每个区域的标签。每张图片选出正/负区域个数相等的共 n 个区域, 那么 $N_{\text{cut_p}}=5$ 时的训练样本矩阵 $P(w|d_{\text{train}})$ 维数就是 $K \times N$, 其中 N 是训练图片提取的区域总个数。对于一张测试样本, 我们使用同样的方法得到 $K \times 5$ 维的区域特征直方图 $P(w|d_{\text{test}})$ 。我们利用 ML-pLSA 模型(如图 4)对训练样本矩阵 $P(w|d_{\text{train}})$ 进行分解, 得到 $K \times Z$ 维矩阵 $P(w|z)$ 和 $Z \times N$ 维矩阵 $P(z|d_{\text{train}})$ 。其中变量 Z 表示潜在主题个数。然后采用迭代操作, 固定 $P(w|z)$ 矩阵, 计算出测试图片的 $Z \times 5$ 维矩阵 $P(z|d_{\text{test}})$ 。最后利用 K 近邻方法对 $P(z|d_{\text{test}})$ 与 $P(z|d_{\text{train}})$ 进行相似性的度量。

接下来, 本文将不同参数下得到的带有标签的图片以每个像素为单位进行叠加, 这样就得到了一

张图片的带有位置信息、类别信息的权值图。最后通过阈值法将置信值高的部分提取出来。

在本文中使用了 3 种评价方法, 一是像素准确率, 即所有判断正确的像素点个数除以图片像素总个数。二是平均像素准确率, 即目标被正确分类的像素个数除以真值像素个数与目标被错误分类的像素个数之和(即 $R \cap G / R \cup G$, 其中 R 表示识别结果区域, G 表示真值区域)。三是像素查准查全率曲线(简称 pr 曲线), 其中精确率(Precision) $p = R \cap G / G$, 回归率(Recall) $r = R \cap G / R$ 。本文将最终得到的权值图量化到 0-255 上, 然后分为 30 层对整个图像库识别结果进行度量。当为 l 层时, 所有大于 $256/l$ 的灰度值都被当做是前景, 其余的为背景, 此时 $p = r$ 的值即为 l 层时 pr 曲线取值。

本文分别做了 3 组实验:

实验 1 以自行车类为例, 先分别用每一种分割方法对测试图片进行实验。然后又使用 ML-pLSA 方法对 150 张测试图片进行实验。4 组数据的平均像素准确率比较见图 6, pr 曲线结果比较见图 7。

从图 6, 图 7 实验数据可以看出, ML-pLSA 算法在测试时的平均像素准确率和 pr 曲线都要好于用单一方法的实验结果。进一步说明多种分割方法的使用是合理的、有效的, 不论单张图片的识别率还是整个图像库的整体表现都要好于使用单一分割方法。

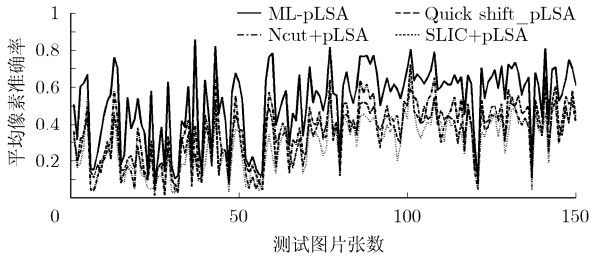


图 6 ML-pLSA 与 X-pLSA 平均像素准确率比较曲线图

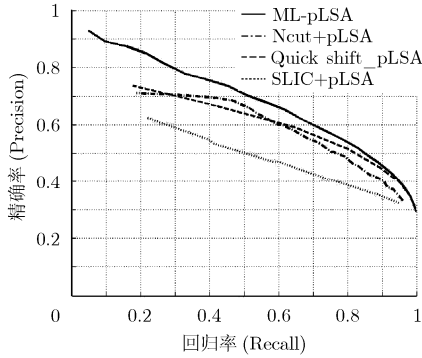


图 7 ML-pLSA 与 X-pLSA pr 曲线比较结果

表 1 列出了实验 1 的相关识别率, ML-pLSA 算法的平均像素准确率和 pr 曲线分别比 X-pLSA 算法 (X 代表单独一种分割方法) 高出 17.25%, 5.04%。

实验 2 ML-pLSA 算法在 GRAZ-02 数据库的结果。图 8 给出了部分测试图片中间结果图。其中

表 1 ML-pLSA 与 X-pLSA 识别率比较(%)

		ML-pLSA	X-pLSA	差值
平均像素准确率	最好	85.56	65.6	19.96
	平均	51.37	34.12	17.25
pr 曲线	150 张图片平均表现	65.12	60.08	5.04

第 1 列为原图, 第 2 列为真值图像, 第 3 至 5 列分别为 Neut, Quick shift, SLIC 分割方法对应的权值图, 第 6 列为本文方法得到的最终结果图。其中原图包含了大目标、小目标、多个目标、角度旋转及光照变化的图片, 对算法具有一定的挑战。但从结果图可见, ML-pLSA 算法的效果还是很理想的。

ML-pLSA 算法在 GRAZ-02 数据库 3 类目标识别率见表 2。从表中可见, ML-pLSA 算法在自行车类中识别效果是最好的。

图 9 为利用 ML-pLSA 算法得到的目标识别结果。

实验 3 ML-pLSA 算法与其他算法的比较

这里将 ML-pLSA 算法分别与 Marszalek^[15]得出的结果和 ECCV2008 Fulkerson^[16]得出的结果进行了比较。这两篇文章评价标准与本文平均像素准确率标准相同, 数据库也是 GRAZ-02 数据库(见表 3)。

结果表明, ML-pLSA 算法在车类和人类的识别

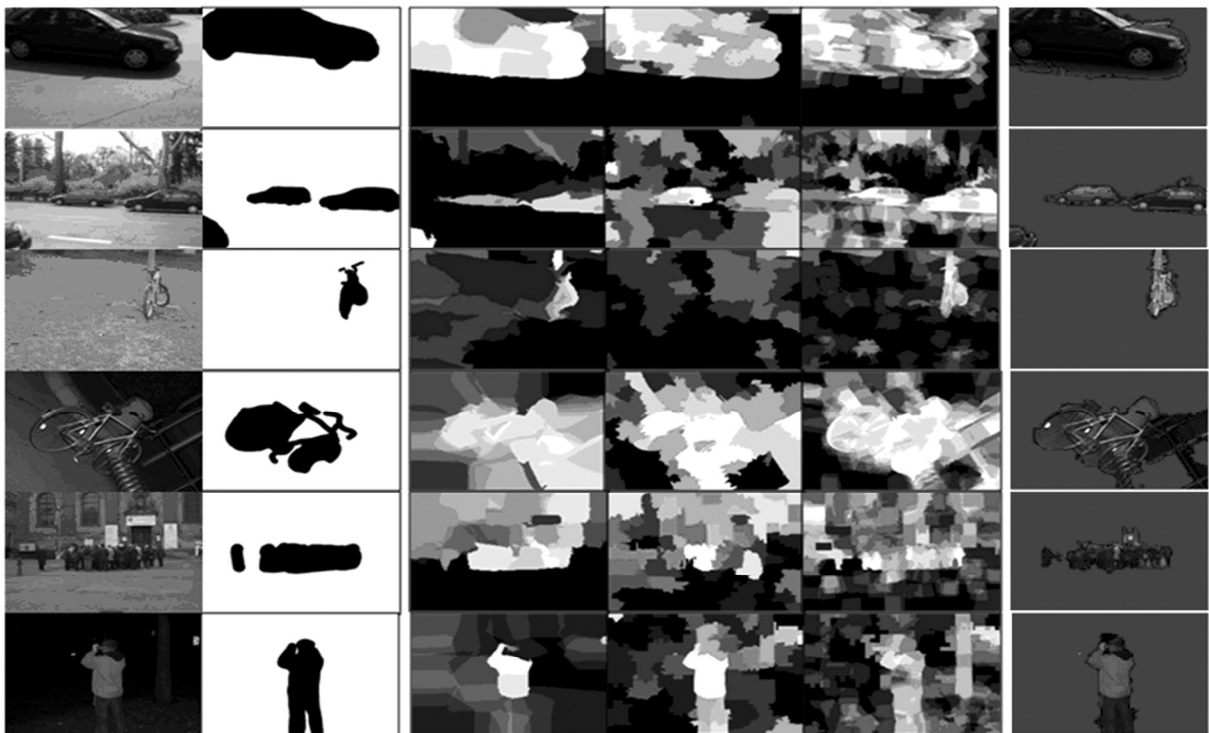


图 8 ML-pLSA 算法在 GRAZ-02 数据库的部分过程示意图

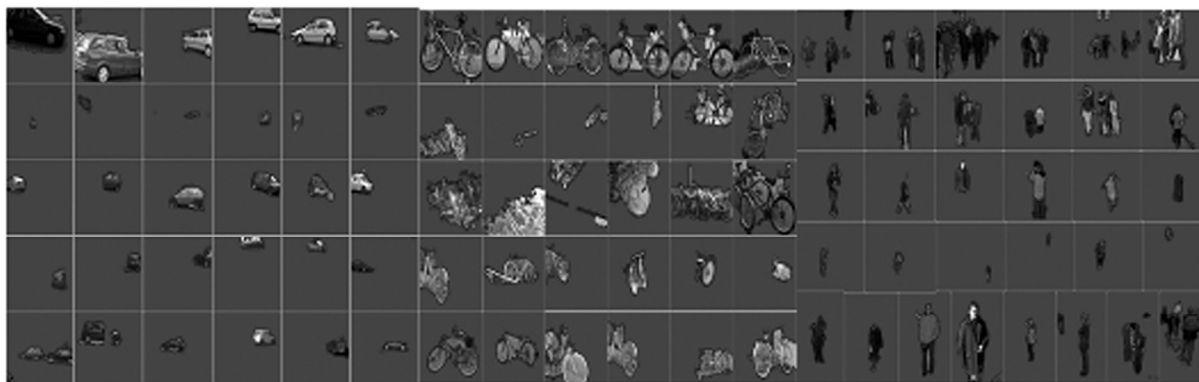


图 9 ML-pLSA 方法结果图

表 2 本方法在图像库中的识别率(%)

评价标准	车	人	自行车
像素准确率	91.22	85.66	82.09
平均像素准确率	45.32	45.85	51.37
pr 曲线	57.19	59.66	65.12

表 3 ML-pLSA 算法与其他算法的比较(%)

算法	车	人	自行车
full framework ^[15]	53.8	44.1	61.8
nearest neighbor ^[16]	54.7	47.1	66.4
SVM ^[16]	49.4	51.4	65.2
ML-pLSA	57.19	59.66	65.12

结果都要远远高于其他目标识别方法; ML-pLSA 算法在自行车类的识别率也接近其他算法的最优值。

5 结束语

本文提出的 ML-pLSA 模型的目标识别算法是一种鲁棒性很强的算法。首先, 相对其他识别算法来说本方法识别率高, 因为它将不同尺度的、丰富的特征空间结合在一起, 更充分地利用了图像的特征信息; 第二, 本方法不局限于先分割再识别的顺序, 而是使整个过程的分割与识别即相互依赖又相互独立, 避免了分割误差对识别过程的影响。第三, 由于使用了多种分割和 SIFT 特征, 所以对目标尺度的变化和光照角度变化也有很好的鲁棒性。本文在 GRAZ-02 数据库做了大量实验, 取得了不错的效果。

在本方法中, 如何提高算法速度、如何更好的融合多特征是本文作者以后要研究的内容。

参考文献

[1] Russell B C, Freeman W T, Alexei A, *et al.* Using multiple segmentations to discover objects and their extent in image

collections [C]. IEEE Conference on Computer Vision and Pattern Recognition, NY, USA, June 17-22, 2006: 1605-1614.

- [2] Gu Chun hui, Lim J J, Arbelaez P, *et al.* Recognition using regions [C]. IEEE Conference on Computer Vision and Pattern Recognition, Florida, USA, June 20-25, 2009: 1030-1037.
- [3] Hofmann T, *et al.* Probabilistic latent semantic analysis [C]. Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, Stockholm, Sweden, July 30-August 1, 1999: 289-296.
- [4] Hofmann T, *et al.* Unsupervised learning by probabilistic latent semantic analysis [J]. *Machine Learning*, 2001, 42 (1/2): 177-196.
- [5] Bosch A, Zisserman A, Munoz X, *et al.* Scene classification via pLSA [C]. European Conference on Computer Vision, Graz, Austria, May 7-13, 2006: 517-530.
- [6] Csurka G, Dance C, Fan L, *et al.* Visual categorization with bags of keypoints [C]. European Conference on Computer Vision, Prague, Czech Republic, March 27, 2004, (1): 1-22.
- [7] Achanta R, Shaji A, Smith K, *et al.* SLIC superpixels[R]. EPFL Technical Report, June 2010.
- [8] Shi Jianbo, Malik J, *et al.* Normalized cuts and image segmentation [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [9] Vedaldi A, Soatto S, *et al.* Quick shift and kernel methods for mode seeking [C]. European Conference on Computer Vision, Marseille, France, October 12-18, 2008: 705-718.
- [10] Levinshtein A, Sminchisescu C, Dickinson S J, *et al.* Optimal contour closure by superpixel grouping [C]. European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010: 480-493.
- [11] Sapp B, Jordan C, Taskar B, *et al.* Adaptive pose priors for pictorial structures [C]. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, June 13-18, 2010: 422-429.

- [12] Fulkerson B, Vedaldi A, Soatto S, *et al.* Class segmentation and object localization with superpixel neighborhoods [C]. International Conference on Computer Vision, Kyoto, Japan, September 27–October 4, 2009: 670–677.
- [13] Ravichandran A, Favaro P, Vidal R, *et al.* A Unified approach to segmentation and categorization of dynamic textures [C]. Asian Conference on Computer Vision, Queenstown, New Zealand, November 8–12, 2010: 425–438.
- [14] Winn J M, Criminisi A, Minka T P, *et al.* Object categorization by learned universal visual dictionary[C]. International Conference on Computer Vision, Beijing, China, October 17–20, 2005: 1800–1807.
- [15] Marszalek M, Schmid C, *et al.* Accurate object localization with shape masks [C]. IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota, USA, June 18–23, 2007: 1–8.
- [16] Fulkerson B, Vedaldi A, Soatto S, *et al.* Localizing objects with smart dictionaries[C]. European Conference on Computer Vision, Marseille, France, October 12–18, 2008: 179–192.
- 陈 琳: 女, 1985 年生, 硕士生, 研究方向为计算机视觉、目标识别.
- 卢湖川: 男, 1972 年生, 博士, 副教授, 博士生导师, 研究方向为图像处理、模式识别、计算机视觉、机器学习、视频跟踪、图像分割.