基于形状统计模型的多类目标自动识别方法

孙　显[①②③]　　王宏琦[①②]　　杨志峰[②]
[①](中国科学院电子学研究所　北京　100190)
[②](中国科学院空间信息处理与应用系统技术重点实验室　北京　100190)
[③](中国科学院研究生院　北京　100190)

摘　要：形状是人类视觉系统分析和识别目标的基础。针对现有方法的不足，该文提出了一种新的基于形状统计模型的多类目标自动识别方法。该模型定义形状基元对作为特征描述子，从样本图像中抽取典型基元对，聚类量化后组成形状字典。然后综合分析各类信息，通过无监督学习来统计目标的特征分布状况，构建类别形状模型。快速定位目标区域并辨识对象类别后，可结合图像分割获取精确形状。实验结果表明，该方法能准确、高效地提取多种类型和复杂结构的目标，较好解决了噪声干扰、旋转侧偏等问题，具有较强的实用价值。

关键词：图像处理；目标识别；形状统计模型；无监督学习
中图分类号：TP391.41　　**文献标识码**：A　　**文章编号**：1009-5896(2009)11-2626-06

Automatic Multi-categorical Objects Recognition Using Shape Statistical Models

Sun Xian[①②③]　　Wang Hong-qi[①②]　　Yang Zhi-feng[②]
[①](*Institute of Electronic, Chinese Academy of Sciences, Beijing* 100190, *China*)
[②](*Key Laboratory of Spatial Information Processing and Application System Technology, Chinese Academy of Sciences, Beijing* 100190, *China*)
[③](*Graduate University, Chinese Academy of Sciences, Beijing* 100190, *China*)

Abstract: Contour features are powerful cues for human vision system to analyze and identify objects. A new method for automatic multi-categorical objects recognition using shape statistical models is proposed to improve the disadvantages existing in most of the relative methods. This method defines firstly the shape base pairs as feature descriptors, and extracts typical shape base pairs from sample images to build a feature codebook. Then, unsupervised learning is performed to calculate the feature distribution and design class-specific shape models. After detecting the regions and determining the categories quickly, segmentation could be applied to obtain the precise outlines. Experimental results demonstrate that proposed method can achieve high efficiency and accuracy in extracting manifold and complicated objects, and resolve the problems of noise disturbance, rotations at a certain extent.

Key words: Image processing; Object recognition; Shape statistical models; Unsupervised learning

1　引言

形状是人类视觉系统感知和分析目标的基础，可以定义为目标的边界或者目标边界所包围的区域。与其它视觉特征如色彩、纹理等相比，形状特征在描述和识别目标中具有较大的优势，包含更丰富的空间语义信息，即使在只具少量外在形状的情况下，人们也可以大致判断目标的类型和结构。目前，形状识别作为计算机视觉和图像处理等领域的一个重要组成部分，已经在工业自动化、多媒体处理和军事情报等方面得到了广泛的应用和发展。

最初的研究方法设计特定的目标外形模板与整幅图像直接匹配进行[1,2]，操作直观形象，但对每类目标都需要一个对应模板，考虑目标形状的多样性，形状模板集和计算开销都很大。部分方法提出对形状模板的几何统计参数进行计算匹配[3]，减轻了工作量，但不适用于复杂形状目标的检测。近年来，有方法尝试从边界片段入手，构建局部到整体的模型[4]来识别目标形状。这类方法在视频跟踪及部分场景目标的识别中得到一定应用，但只对平滑、规则的边界提取效果较好，且需要大量的人工交互。Opelt[5]和 Shotton[6]等在此基础上做了改进，通过对样本进

2008-11-03 收到，2009-06-25 改回
国家自然科学基金(40871209)，国家 863 计划项目(2006AA12Z149)和中国科学院电子学研究所青年创新基金资助课题

行标记学习，提高了处理效率。但由于对形状特性的描述尚不完善，获取的信息较有限，在面对复杂目标或背景噪声干扰大的图像时，提取结果不够理想。

针对以上不足，本文提出了一种新的基于形状统计模型的多类目标自动识别方法。该模型定义形状基元对作为目标特征描述子，构建相应的形状字典，通过无监督学习来构建符合各个类别的形状模型，可在图像中准确识别多种类型和复杂结构的目标对象，有效消除了绝大部分噪声及冗余背景的干扰，智能化程度高、鲁棒性好，具有较强的稳定性和实用性。

2　研究方法

基于形状统计模型的多类目标识别方法流程大致分为 4 个部分：首先提取形状基元对作为目标特征描述子；其次筛选基元对组成形状字典；然后分析隐含概率语义，构建各个类别的形状语义模型；最后对目标检测并获得识别结果。下面对各个部分进行详细阐述。

2.1 形状基元对提取

任意一段从目标外部或内部边界截取的曲线均可视为形状基元。本文中将训练数据划分为提取图像集和评估图像集两类，前者包含待识别的目标，用于提供候选形状基元，后者用于评估候选基元的有效性，分为包含目标的正样本图像和不含目标的负样本图像。经过 Canny 算子处理后，用编组技巧连接有效短边[7]，滤除边界噪声。

本文提出的方法中将形状基元两两配对，以形状基元对作为特征描述子。相比单个基元，形状基元对受到更严格的空间几何限制，能更好地体现目标形状特性。定义形状基元对 (i,j) 为：$r_{ij}=(l_i,l_j,s_i,s_j,\theta_i,\theta_j,\varphi_{ij},\varphi_{ji},d_{ij})$。如图 1 所示，$l$ 为形状基元的像素列表，s 为尺度因子，令 $s=\sqrt{A(b)}$，b 为形状基元对应的外接矩形，A 为 b 的面积，θ 为基元法线与水平线间的夹角，φ 为法线与基元中点连线的夹角，d 为中点间的距离。依此得到所有基元对的集合 $R=\{r_{11},\cdots,r_{1n},\cdots,r_{mn}\}$。每个基元对在保持基本信息不变的情况下，都是可以自由变化的。

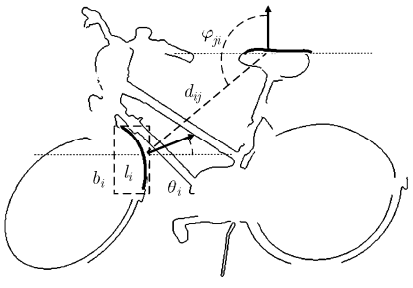

图 1 形状基元对示意图

$$r_{ij}^{'}=r_{ij},\quad i=1,2,\cdots,m,\quad j=1,2,\cdots,n \tag{1}$$

式中 $r_{ij}\in R$，$r_{ij}^{'}$ 为 r_{ij} 的变换形式。当且仅当存在 (λ_t,δ_t) 满足以下条件时，等式成立：

$$\left.\begin{aligned}
&l_i^{'}=\mathrm{trans}(l_i,\lambda_t),\ l_j^{'}=\mathrm{trans}(l_j,\lambda_t)\\
&s_i^{'}=\lambda_t s_i,\ s_j^{'}=\lambda_t s_j\\
&\theta_i^{'}=\theta_i+\delta_t,\ \theta_j^{'}=\theta_j+\delta_t\\
&\varphi_{ij}^{'}=\varphi_{ij},\ \varphi_{ij}^{'}=\varphi_{ji}\\
&d_{ij}^{'}=\lambda_t d_{ij}\\
&\lambda_t\in[-x_1,x_1],\ \delta_t\in[-x_2,x_2]
\end{aligned}\right\} \tag{2}$$

其中 λ_t 为比例因子，δ_t 为旋转因子，$\mathrm{trans}(\bullet)$ 为曲线段按比例 λ_t 的最近邻域插值变形。x_1,x_2 分别为 λ_t,δ_t 的取值上下限，例如可令 $x_1=10$，$x_2=\pi/15$。

可见，集合 R 中包含的形状基元对，不但可以经过自主变换来组合表达目标图像的形状信息，从而摆脱了尺度缩放、角度旋转、视角转换等因素带来的制约，而且由于基元对的精度可以达到像素级别，且具备丰富的空间信息，提升了目标区域定位的准确度。

2.2 形状字典构建

度量筛选获取的大量形状基元对，量化聚类后组成形状字典。该字典能够完整涵盖所有类别的特征，成为下一步评估目标类别的统计标准。形状字典的构建步骤如下：

(1)筛选典型基元对。假设 R' 为基元对 (i,j) 在给定空间 $\Delta=[-x_1,x_1]\times[-x_2,x_2]$ 中得到的变换基元集合，本文使用 C_{ij} 来度量形状基元对及其变换形式的表达质量好坏：

$$C_{ij}=\min_{(i,j)\subset R'}\frac{\frac{1}{N^+}\sum_{i,j,a^+,b^+=1}^{N^+}G_{ia^+;jb^+}}{\frac{1}{N^-}\sum_{i,j,a^-,b^-=1}^{N^-}G_{ia^-;jb^-}} \tag{3}$$

其中 N^+ 和 N^- 分别代表正负样本图片 I^+ 和 I^- 的数量，$G_{ia;jb}$ 为条件随机场对数分类器[8]，可将度量基元对 (i,j) 和 (a,b) 间几何相似关系的向量 $\boldsymbol{g}_{ij}(a,b)$ 映射为类概率的形式：

$$G_{ia;jb}=\frac{1}{1+\exp(-\boldsymbol{w}^{\mathrm{T}}\boldsymbol{g}_{ij}(a^+,b^+))} \tag{4}$$

其中 $\boldsymbol{w}^{\mathrm{T}}$ 为模型参数，$\boldsymbol{g}_{ij}(a,b)=[1,e(1)^2,\cdots,e(9)^2]$，有

$$e(h)=\begin{cases}\mathrm{DT}[r_{ij}(h),r_{ab}(h)], & h=1,2\\ r_{ij}(h)-r_{ab}(h), & h=3,\cdots,9\end{cases} \tag{5}$$

$\mathrm{DT}(\bullet)$ 表示 Chamfer 距离变换[9]，它可以较准确地

度量两段曲线间的相似度。

遍历评估图像集筛选基元对。C_{ij} 越小，说明对正负样本图像的区分度越好，越有可能是目标的有效特征描述子即典型基元对，应当保留，反之则丢弃。实际中，为了减少计算量，将每幅样本图像中的基元对按 C_{ij} 由小到大排序，取匹配最优的 20 个基元对用于统计。

(2)确定聚类中心，量化典型基元对。为选取最优的聚类中心数目，引用文献[10]中的方法，先用高斯混合模型对所有筛选得到的形状基元对建模，然后结合最小描述长度准则估计模型复杂度，最优复杂度对应的值即为聚类中心的个数，假定为 M 。然后利用 k-means 算法聚类，将每个形状基元对按照聚类中心进行量化。一方面将相似形状基元对归类，去除随机抽取时产生的重复元素，降低字典冗余度，另一方面保证特征描述的稳定性，避免出现畸变误差。最终选出的形状基元对构成了形状基元字典 D 。图 2 显示了部分典型基元对及其聚类中心。实验中为保证算法效率，形状字典的总量尽量以不超过 100 个为宜。

2.3 统计学习

构建形状字典后，可以将目标特征信息用形状基元对重新表达。本文采用隐含概率语义分析(probabilistic latent semantic analysis)算法[11]，通过无监督的统计学习来分析图像中包含的形状等可视及隐含特性、关联语义等非可视化信息，构建各个类别的形状模型。

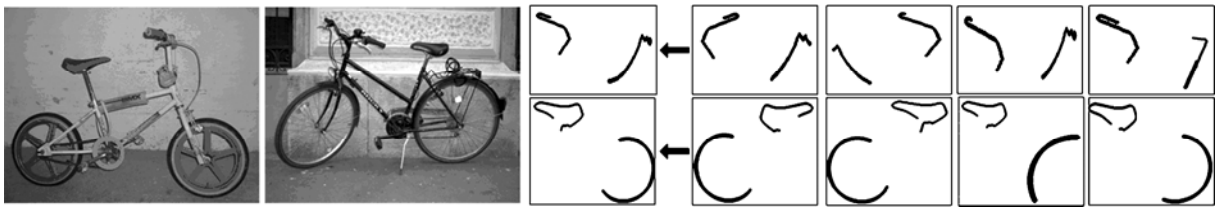

图 2 形状基元字典构建

隐含概率语义分析算法最初应用于文本检索。引入图像分析中时，将图像对象对应为文档，将目标属于的类别对应为文档主题，而字典中的形状基元对则是文档中出现的单词。从形状字典 D 中选取典型基元对重新表达所有的评估图像，统计每幅图像中这些形状基元对的分布状况，而后将这些分布拟合成若干个已知主题类别的混合直方图形式。

假设存在 U 个样本图像，对应 V 个特征分布，$n(k_i,\ f_j)$ 表示图像 k_i 中特征基元对 f_j 的出现次数，而隐含目标类别变量 t_q 与单个基元对在特定对象中的出现概率相关联。隐含概率语义分析算法对图像对象、类别和特征这三者进行统计建模 $P(f_j,\ k_i,\ t_q)$，如图 3 所示，阴影结点为可观察到的随机变量，非阴影结点为不可观察到的变量信息。可以发现，特征和对象的产生式推断 $P(k_i,\ f_j)=P(k_i)P(f_j\,|k_i)$ 能够通过条件概率计算得到：

$$P(f_j\mid k_i)=\sum_{q=1}^{Q}P(f_j\,|t_q)P(t_q\,|k_i) \tag{6}$$

其中条件分布 $P(f_j\,|t_q)$ 建模特征和隐含目标类别间的关系，$P(t_q\,|k_i)$ 建模目标类别和对象之间的关系，Q 为目标类别的数目。

隐含概率语义分析算法的目的是学习概率 $P(f_j\,|t_q)$ 和 $P(t_q\,|k_i)$，这些概率可以通过期望最大化(EM)[12]算法拟合得到：

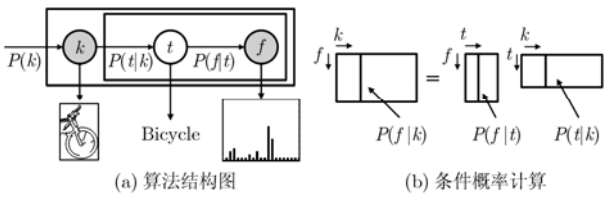

图 3 隐含概率语义分析算法

$$\text{E步骤 } P(t_q|k_i,f_j)=\frac{P(f_j|t_q)P(t_q\,|k_i)}{\sum_{q=1}^{Q}P(f_j|t_q)P(t_q|k_i)} \tag{7}$$

M步骤

$$\begin{cases}P(f_j\,|t_k)=\dfrac{\sum_{i=1}^{V}n(k_i,f_j)P(t_q|k_i,f_j)}{\sum_{u=1}^{U}\sum_{i=1}^{V}n(k_i,f_u)P(t_q\,|k_i,f_u)}\\[2ex] P(t_q\,|k_i)=\dfrac{\sum_{j=1}^{U}n(k_i,f_j)P(t_q|k_i,f_j)}{\sum_{j=1}^{U}n(k_i,f_j)}\end{cases} \tag{8}$$

E 步骤和 M 步骤重复迭代直至满足期望约束。求得各个参数后，即可完成各个形状统计模型的构建。在模型计算后，可以得到目标的潜在语义空间，该潜在因素量化了目标包含的个性化信息，能够表达出目标类别区别于其它类别的语义关联，以此快速、有效地将目标与背景噪声及非目标类别区分开来。

2.4 目标识别

本文方法利用可变滑动窗口来识别多类目标。

如图 4 所示，对待检测图像先做边缘提取，滑动窗扫描整幅图像，对窗口内的边界部分稀疏采样，提取形状基元对后，根据形状字典生成其表示 $n'(k_i, f_j)$。接下来可计算特征分布并按照式(6)–式(8)求解对应的标记概率 $P(t_q|k_i)$。标记概率值越大，说明窗口内图像属于 q 类目标的可能性越大，反之亦然，令 η_d 为识别阈值，若 $P(t_q|k_i)>\eta_d$，则判断属于目标类，标记出检测框和类别。为提高检测精度，还需对窗口进行多尺度缩放。假设最小尺度为 1.0，缩放的尺度因子为 ρ，尺度的数量可以根据实际图像的情况进行调整，滑动窗每次移动的距离为 $dx=1.0\times\rho$。对于目标附近可能出现的多个检测框，以检测框之间的中心距离及相互覆盖程度为准则进行合并。这样，不但可以有效避免背景噪声的干扰，而且可以得到具有较高置信度的目标区域。用图像分割算法如 Log-cut[13]等做进一步精处理，可以提取目标的准确形状。

3　实验分析

为验证本文提出方法的有效性，本节中对两类图像数据进行了实验。一类为 Graz17 数据集，由包含 17 类不同目标的自然场景图像组成[5]，其中部分目标如牛、汽车等按照不同的视角还单独划为几类，分别选取与文献[5,6]等相同的图像，实验中截取图像中包含目标的区域作为修正后的正样本训练集。另一类为 HRRSI 数据集，由从因特网上搜集到的高分辨率遥感图像(high resolution remote sensing imagery)组成，图像分辨率为 1 m 左右，包含舰船、油罐、飞机 3 类地物目标，每类目标均有 500 幅图像，图像的平均尺寸约为 400×280 像素大小。对每一类目标按照 1:4 的比例选取训练和测试图像。相比自然场景图像，这类目标形状复杂，容易受到周围背景遮挡、边界混杂等干扰，提取更为困难。

本文从两方面来评估识别算法的性能：首先通过观察检测外接矩形 r_{pred} 与实际目标外接矩形 r_{gt} 是否满足 $\dfrac{\text{aera}(r_{\text{pred}}\cap r_{gt})}{\text{aera}(r_{\text{pred}}\cup r_{gt})}>0.5$ 来判断结果是否准确[6]，每个目标只对应一个检测矩形；其次，定义 RPC 曲线和平均等错误率(EER)等指标来评价数据集测试结果。RPC 曲线是显示分类模型检测率(Recall)和准确率(Precision)之间折中的一种图形化表达方法，越好的模型越靠近图形左上角，其曲线下方面积(AUC)就相对越大。平均等错误率(EER) 则用于评价模型的平均性能，可通过调节阈值使误识率(FAR)和拒识率(FRR)两个指标相等时得到。

图 5 显示了在相同样本和参数条件下，采用一个或多个(两个以上)形状基元作为模型组成单元进行实验后得到的 RPC 曲线。单个基元由于仅利用了边界匹配信息，存在一定局限性；而组合的基元数目过多，固然可以增加信息量，但也会增加计算负担，并容易造成漏检。因此，采用基元对作为组成形状统计模型的基本单元是一个较为合理的选择。

本文方法通过统计学习来构建模型。图 6 给出了不同样本数目对训练学习产生的影响。可以发现，

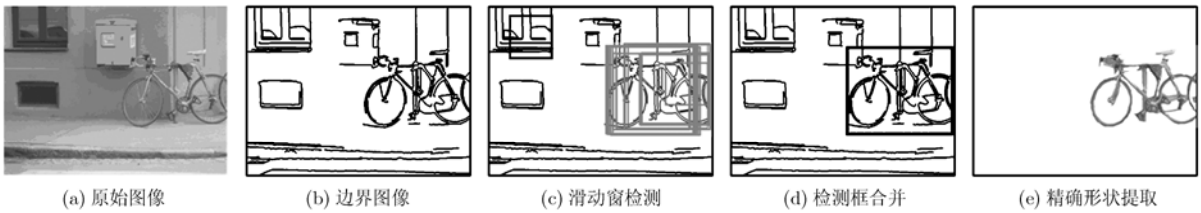

图 4　目标识别流程

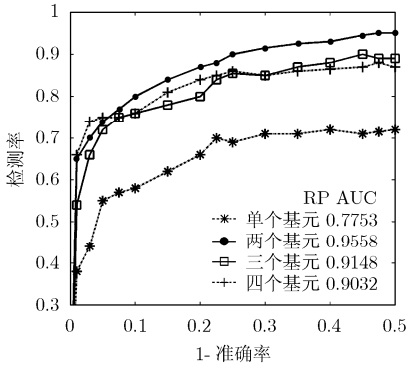

图 5 不同数目基元组成形状统计模型的 RPC 曲线

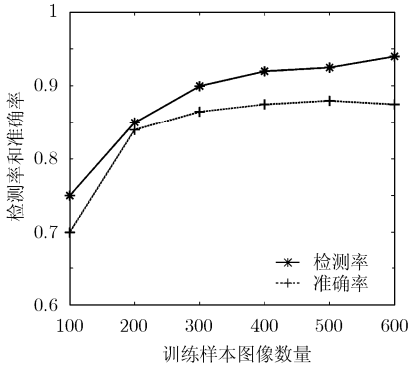

图 6　不同样本数目对训练学习的影响

只需少量样本，方法的识别性能就可以显著提高，并迅速达到稳定。

图 7 为部分测试图像及识别结果，前两幅为 Graz17 数据集图像，后两幅为 HRRSI 数据集图像。可以发现，图中部分目标，如舰船、油罐等，虽然由于图像来源或光照影响不同，目标的色彩、纹理、形态各异，但本文方法通过构建形状统计模型，仍然能够正确地将其归为一类。对于部分被局部遮挡或存在一定程度旋转、侧偏、尺度变化的目标，如自行车等，该方法也能很好地定位目标区域。此外，本方法还较好的解决了图像中背景噪声较多、目标边界模糊等带来的问题，能够快速排除干扰，具有较高的识别性能。

表 1 给出了本文方法识别 Graz17 数据集中目标的性能指标，并与文献[5,6]中的结果相比较。对比指标为检测平均等错误率和分类平均等错误率两类。由于文献[5]只给出了检测时的平均等错误率，分类指标有所缺失。从表中可以看出，本文方法对绝大多数类别都具有较好的识别度，特别是在一些形状多变、背景较为复杂的目标，如自行车、汽车、杯子等上有较大改进。由于部分目标类别提供的图像数量较少，不同图像间的目标差异明显，导致形状模型构建不够全面，性能指标出现波动。如果适度增加样本图像的数量，这种情况可以得到改善。总体而言，基于形状统计模型的多类目标识别方法是有效实用的。

4 结束语

本文提出了一种基于形状统计模型来识别多类目标的方法。该方法在兼具现有方法优点的同时，以形状基元对为基础构建统计模型，能不受尺度、旋转等变化的影响，准确描述各类目标的结构信息。考虑图像中存在大量干扰噪声和背景线群，本文方法充分利用了目标对象中包含的各类可视化和非可视化信息，通过无监督学习来统计典型基元对的分布状况，并整体分析隐含概率语义，可以在图像中快速辨识对象类别并定位目标位置区域。

实验结果表明，本文提出的方法适用于多种类型和复杂结构目标的提取，具有较强的实用价值。同时，该方法在一定程度上也表明了形状特征在目标检测与识别中所起的有效作用，如何对其进行扩展，或者结合其它特征进一步提高识别率，是今后值得继续研究的课题。

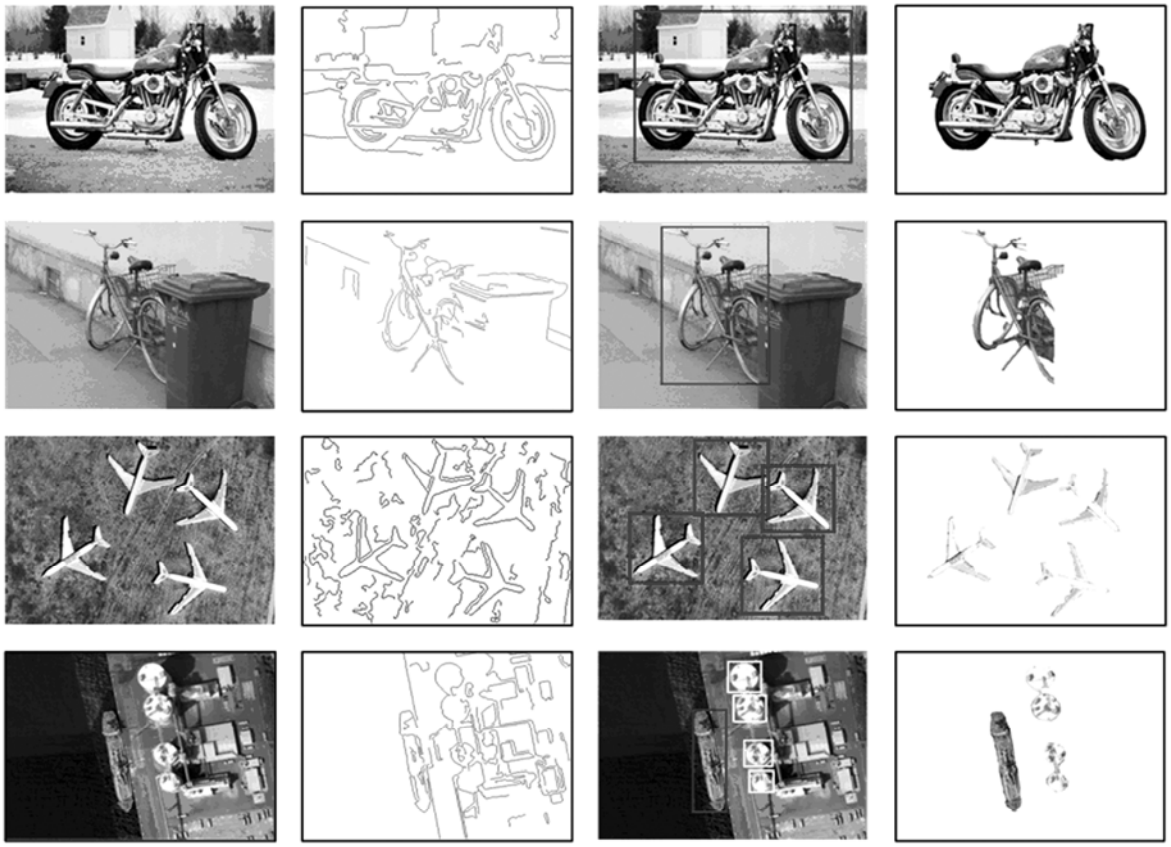

图 7 部分测试图像及识别结果

表 1 本文方法与其它部分方法识别 Graz17 数据集的性能比较

	图像数量		检测(%)			分类(%)	
类别	训练图像	测试图像	**EER**(本文)	EER[5]	EER[6]	**EER**(本文)	EER[6]
Airplanes	50	200	**5.0**	7.4	6.8	**2.5**	3.4
Cars(rear)	50	200	**1.5**	2.3	1.8	**1.3**	1.5
Motorbikes	50	200	**0.3**	4.4	0.3	**0.3**	0.4
Faces	50	100	**3.3**	3.6	2.8	**2.3**	2.4
Bikes(side)	45	53	**17.0**	28.0	32.1	**9.4**	13.2
Bikes(rear)	29	13	**23.1**	25.0	26.7	**7.7**	15.4
Bikes(front)	19	12	**16.7**	41.7	41.7	**8.3**	16.7
Cars(2/3rear)	32	14	**7.1**	12.5	30.0	**7.1**	20.9
Cars(front)	34	16	**6.3**	10.0	29.4	**6.3**	12.5
Bottles	54	64	**4.7**	9.0	9.4	**3.1**	7.8
Cows(side)	45	65	**1.5**	0.0	1.5	**1.5**	1.7
Horses(side)	55	96	**7.3**	8.2	6.3	**5.2**	6.3
Horses(font)	44	22	**13.6**	13.8	27.3	**9.1**	13.6
Cows(front)	34	16	**12.5**	18.0	18.8	**6.3**	6.3
People	39	18	**38.4**	47.4	47.6	**11.1**	16.7
Mugs	30	20	**5.0**	6.7	10.0	**5.0**	5.0
Cups	31	20	**10.0**	18.8	15.0	**5.0**	5.0

参 考 文 献

[1] Leibe B, Leonardis A, and Schiele B. Robust object detection with interleaved categorization and segmentation [J]. *International Journal of Computer Vision Special Issue on Learning for Recognition and Recognition for Learning*, 2008, 77(1): 259–289.

[2] Felzenszwalb P F and Schwartz J D. Hierarchical matching of deformable shapes [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minnesota, USA, 2007: 1–8.

[3] Zhang X Q, Guo M M, and Tang Y. A new geometric feature shape descriptor [J]. *Computer Engineering and Applications*, 2007, 43(29): 90–92.

[4] Fergus R, Perona P, and Zisserman A. Weakly supervised scale-invariant learning of models for visual recognition [J]. *International Journal of Computer Vision*, 2007, 71(3): 273–303.

[5] Opelt A, Pinz A, and Zisserman A. Incremental learning of object detectors using a visual shape alphabet [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 3–10.

[6] Shotton J, Blake A, and Cipolla R. Multi-scale categorical object recognition using contour fragments [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(7): 1270–1281.

[7] Mahamud S, Williams L R, Thornber K, and Xu K. Segmentation of multiple salient closed contours from real images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(4): 433–444.

[8] Kumar S. Models for learning spatial interactions in natural images for context-based classification [D]. [Ph.D. dissertation], The Robotics Institute, Carnegie Mellon University, 2005.

[9] Wu Tao, Ding Xiao-qing, and Wang Sheng-jin. Video tracking using improved chamfer matching and particle filter [C]. Proceedings of IEEE Conference on Computational Intelligence and Multimedia Applications, Sivakasi, Tamil Nadu, 2007: 169–173.

[10] Maitre H, Kyrgyzov I, and Campedel M. Kernel mdl to determine the number of clusters [C]. Proceedings of IEEE Conference on Machine Learning and Data Mining in Pattern Recognition, Leipzig, Germany, 2007: 203–217.

[11] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis [J]. *Machine Learning*, 2001, 42(2): 177–196.

[12] Dempster A, Laird N, and Rubin D. Maximum likelihood form incomplete data via the EM algorithm [J]. *Journal of the Royal Statistical Scotiety*, 1977, 39(1): 1–38.

[13] Lempitsky V, Rother C, and Blake A. LogCut-efficient graph cut optimization for markov random fields [C]. Proceedings of IEEE Conference on Computer Vision, Rio de Janeiro, Brazil, 2007: 1–8.

孙　显：　男，1981 年生，博士生，研究方向为遥感图像处理、人工智能.

王宏琦：　男，1964 年生，研究员，博士生导师，研究方向为信号与信息处理、图像处理.

杨志峰：　男，1982 年生，硕士生，研究方向为图像处理、软件工程.