

文章编号: 1002-0446(2005)02-0123-05

基于场景识别的移动机器人定位方法研究*

李桂芝¹, 安成万², 杨国胜², 谭民², 涂序彦¹

(1. 北京科技大学信息学院, 北京 100083; 2. 中国科学院自动化研究所复杂系统与智能科学开放实验室, 北京 100080)

摘要: 提出了一种基于场景识别的移动机器人定位方法. 对 CCD 采集的工作环境的系列场景图像, 用多通道 Gabor 滤波器提取场景图像的全局纹理特征, 然后通过 SVM 分类器来识别场景图像, 实现机器人的逻辑定位. 在移动机器人 CASIA-I 上对该算法进行了实验. 实验结果表明, 该定位方法可达到 91.11% 的定位准确率, 对光照、对比度等因素有较强的鲁棒性, 并且满足机器人实时定位的要求.

关键词: 移动机器人; 定位; Gabor 滤波器; SVM

中图分类号: TP24 **文献标识码:** B

Scene Recognition for Mobile Robot Localization

LI Gui-zhi¹, AN Cheng-wan², YANG Guo-sheng², TAN Min², TU Xu-yan¹

(1. Institute of Information, University of Science and Technology Beijing, Beijing 100083, China;

2. Laboratory of Complex Systems and Intelligence Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: This paper proposes a scene recognition approach for mobile robot localization. The multi-channel Gabor filters are used to extract the global texture features of the scene images which are associated with the corresponding locations, and then these texture features are fed back to support the vector machine classifier to determine the logical location of the robot. The algorithm has been tested on the autonomous mobile robot CASIA-I designed and developed by us. The experiment results indicate that the algorithm can reach up to a correct localization rate of 91.11%, is robust to the various illumination and contrast, and satisfies the real-time localization demand of the mobile robot.

Keywords: mobile robot; localization; Gabor filter; SVM

1 引言 (Introduction)

移动机器人的视觉导航主要包括定位、路径规划和避障等几方面, 其中定位是移动机器人导航中最基本的环节. 机器人的工作环境、配备传感器的种类和数量的不同使得移动机器人有多种定位方法, 而基于路标的定位方法和基于图像匹配的场景识别定位方法是目前最常用的两种定位方法.

路标可分为人工路标 (如: 直线、箭头或其他图形) 和自然路标 (如: 门、窗户、拐角、灯、柱子). 所谓路标定位是从图像中分割、识别出路标, 根据这些路标在环境中的已知坐标、形状等特征来确定机器人的相对位置. 因此, 这些路标应具有明显的特征并且应很容易从场景图像中被分割出来. 但是在非结构化的环境中, 当不能设置人工路标或者自然路标很难分割时, 使用全局的场景特征来进行定位是非常

有效的方法之一.

基于全局场景特征的定位方法不需要对图像进行分割和局部特征提取, 而是用全局特征识别场景图像, 进而确定机器人的位置. 这种定位方法是一种逻辑定位, 并不给出机器人位置的确切坐标, 而是确定机器人在完成任务过程中所处的场景位置. 基于图像全局特征的场景识别定位方法分为训练与识别定位两个阶段. 在训练阶段, 机器人采集工作环境的图像, 这些图像分别与相应的场景关联, 然后对图像进行处理, 提取用于场景匹配的全局特征, 建立场景图像的样本数据库. 在识别定位阶段, 对机器人所采集的场景图像, 在特征空间中搜索与之最匹配的场景类别, 进而确定机器人在工作环境中的位置^[1,2].

由此可以看出, 基于全局特征的场景识别定位方法关键在于场景全局特征的获取和这些特征的匹

* 基金项目: 国家 863 计划机器人技术主题资助项目 (2002AA423160).

收稿日期: 2004 - 07 - 06

配,这是图像检索在机器人技术中的应用,即对机器人实时采集的序列图像的分类、识别.在图像检索中,广泛使用 Gabor 滤波器提取图像的纹理特征^[3,4],这是因为 Gabor 滤波器提取图像的纹理特征有着良好的视觉特性和生物学背景^[5],而且使用 Gabor 滤波器提取图像的纹理特征对光照、对比度等因素具有很强的鲁棒性.文献[6]指出,在多种小波变换形式的图像纹理特征提取方法中,Gabor 小波变换的图像检索效果最好.也正因为如此,本文拟采用 Gabor 滤波器来提取场景全局特征.

基于全局场景特征的定位方法的另一个关键问题是特征匹配,这属于模式分类的范畴.对于模式分类常用的方法有:神经网络、贝叶斯法、模糊 C 均值(FCM)、近邻法、支持向量机(SVM)等.其中 SVM 是 Vapnik 等人根据统计学习理论提出的一种新的机器学习方法,已经在模式识别、回归分析和特征选择等方面得到了较好的应用效果^[7,8].SVM 与传统方法相比具有运算速度快、分类效果好等优点,特别是对小样本、非线性模式识别问题有很好的效用.由于机器人采集的场景样本图像非常有限,属于小样本识别问题.鉴于此,本文采用 SVM 来解决基于全局特征的场景匹配问题.

2 场景全局特征提取(Global scene features extraction)

2.1 Gabor 滤波器

Gabor 滤波器具有可同时在频率和相位上选择特征的特性,因此,利用 Gabor 滤波器可以提取图像的多个全局特征.多通道 Gabor 滤波器已经被证明是分析具有纹理特性的图像的有效工具^[9],本文采用具有正交相位关系的各向同性 Gabor 滤波器^[10].它的时域数学模型如下:

$$h_e(x, y) = g(x, y) \cos[2\pi f(x \cos \theta + y \sin \theta)] \quad (1)$$

$$h_o(x, y) = g(x, y) \sin[2\pi f(x \cos \theta + y \sin \theta)] \quad (2)$$

$$g(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right] \quad (3)$$

其中 $h_e(x, y)$ 表示偶对称 Gabor 滤波器, $h_o(x, y)$ 表示奇对称 Gabor 滤波器, $g(x, y)$ 是高斯函数, f 为滤波器的径向中心频率, θ 为方向, σ 为空间常数.

2.2 特征提取

每一对 Gabor 滤波器 $h_e(x, y)$ 、 $h_o(x, y)$ 对应于一个特定的中心频率 f 和方向 θ .实验证明:就纹理识别而言,没有必要选择覆盖整个频率域的滤波器空间参数^[10].由于所使用的 Gabor 滤波器是在频率域

中心对称的,因此方向 θ 的选取只需考虑半个频率平面,一般情况下,选择方向 $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$.实践表明,对大多数纹理特征的提取,以上 4 种方向基本能满足分类要求.另外,对于一幅大小为 $N \times N$ 的图像而言,中心频率 f 的选择一般满足 $f \leq N/4$ 的条件.在实验中,图像全部归一化到 128×128 像素,所以对于每个方向 θ ,中心频率 f 选为 2、4、8、16、32.这样得到 20 个 Gabor 滤波器通道.决定每个通道带宽的空间常数 σ 的选择与通道的中心频率成反比.

对于一幅图像,利用 Gabor 滤波器提取的纹理特征是图像经 Gabor 滤波器滤波后输出图像的均值和方差.对于输入图像 $p(x, y)$,经过 Gabor 滤波器滤波后的输出图像为:

$$q(x, y) = \sqrt{[h_e(x, y) \otimes p(x, y)]^2 + [h_o(x, y) \otimes p(x, y)]^2} \quad (4)$$

此处 \otimes 表示卷积,则该通道滤波器提取的图像纹理特征为输出图像 $q(x, y)$ 的均值 E 与方差 D , 其式为:

$$E = \frac{\sum_{x=1}^N \sum_{y=1}^N q(x, y)}{N^2} \quad (5)$$

$$D = \sqrt{\frac{\sum_{x=1}^N \sum_{y=1}^N [q(x, y) - E]^2}{N^2}} \quad (6)$$

这样对于机器人采集的图像,经预处理归一化为 128×128 像素,之后用上述 20 个通道 Gabor 滤波器对图像在不同空间频率和方向上进行卷积,提取出该图像的 40 个特征,作为场景识别分类器的输入样本数据.

3 基于 SVM 的场景识别(Scene recognition based on SVM)

3.1 SVM 学习^[7]

SVM 算法的基本思想是 Vapnik 提出的结构风险最小化原理,通过最大化分类间隔来提高学习机的泛化性能.它通过选择训练一组称为支持向量的特征子集,使得对支持向量集的线性划分等价于对整个数据集的分割,在保证分类精度的同时降低了运算的复杂度.

3.1.1 线性可分情况

设 $\Omega = \{(x_i, y_i) | i = 1, 2, \dots, N\} \subset R^m \times \{-1, +1\}$ 为训练样本集,其中 $x_i \in R^m, y_i \in \{-1, +1\}$.设 Ω 是线性可分的,即存在超平面 $\omega x + b = 0, x \in R^m$,使得对于任一 $(x_i, y_i) \in \Omega$,满足:

$$y_i (\omega x_i + b) \geq 1 \quad i = 1, 2, \dots, N \quad (7)$$

此时, 两类样本到超平面的最小的距离之和即分类间隔为 $2/\|\omega\|$, 而最优分类面是使分类间隔最大, 即使 $\|\omega\|$ 取最小值的分类面. 其中, 使式 (7) 等号成立的样本点称作支持向量. 求解最优分类面可以表示成如下约束的优化问题: 在式 (7) 的约束下, 求解函数 $\phi(\omega) = \|\omega\|^2/2$ 的最小值. 解这个优化问题便得到最优分类函数:

$$f(x) = \operatorname{sgn} \left| \sum_{i=1}^N \alpha_i^* y_i (x, x_i) + b^* \right| \quad (8)$$

式中: $\operatorname{sgn}(\cdot)$ 是符号函数, α_i^* 为 Lagrange 系数, b^* 为分类器阈值. 在样本数据中, 非支持向量对应的 α_i^* 均为 0, 因此上式的求和实际上只对支持向量进行.

3.1.2 非线性可分情况

在训练样本非线性可分的情况下, 可以通过非线性变换 $\phi(x)$ 把训练样本变换到高维特征空间, 使其在高维空间内是线性可分的, 再在高维特征空间中求最优分类超平面. 在线性可分情况下, 最优分类函数中只包含待分类样本和训练样本中支持向量的点积运算 (x, x_i) , 那么要解决高维特征空间中的最优线性分类问题, 只需知道这个空间中的点积运算 $(\phi(x) \cdot \phi(x_i))$ 即可. 因此, 可以利用核函数技术, 以避免向高维特征空间的非线性变换: 即用核函数 $k(x \cdot x_i)$ 代替非线性变换特征空间中的点积. 这就相当于把原特征空间变换到一个新的特征空间, 此时最优分类函数变为:

$$f(x) = \operatorname{sgn} \left| \sum_{i=1}^N \alpha_i^* y_i k(x, x_i) + b^* \right| \quad (9)$$

通常, 采用的核函数有以下 3 类:

(1) 多项式核函数

$$k(x, x_i) = [(x \cdot x_i) + 1]^d \quad (10)$$

(2) 径向基核函数 (RBF)

$$k(x, x_i) = \exp\{-\delta \|x - x_i\|^2\} \quad (11)$$

(3) Sigmoid 核函数

$$k(x, x_i) = \tanh[v(x \cdot x_i) + c] \quad (12)$$

其中, d, δ, v 与 c 分别为多项式核函数、RBF 核函数和 Sigmoid 核函数的调节参数. 采用不同的核函数, 可以生成不同的支持向量机, 从而在特征空间中对应着不同的最优分类超平面, 目前主要依靠经验或实验来选择核函数, 在理论上尚无直接选择法.

3.2 场景识别

SVM 算法最初是为解决二类分类问题而设计的, 当处理多类分类问题时, 需要构造合适的多类分类器. 多类分类器可以通过组合多个二类分类器来

实现, 通常有“一对多”^[11]与“一对一”^[12]两种策略. 对于包含 n 类数据的训练样本, “一对多”策略是在每一类与其余的各类之间构造二类分类器, 因此需要构造 n 个分类器, 每个样本均需要参与 n 次训练; “一对一”策略是在任意两类之间构造二类分类器, 因此需要构造 C_n^2 个分类器, 每个样本参与训练的平均次数为 $(n-1)/2$. 这两种策略各有不同的优势: 前者的分类器数目少, 但样本的训练次数多; 后者的分类器数目多, 但样本的训练次数较少. 但是, 对于一个测试样本, 前者需要 n 次分类, 后者需要 C_n^2 次分类, 所以前者的测试时间要远远小于后者. 通过比较发现, 这两类算法的分类准确度接近^[13].

考虑到机器人在线运行实时性的要求, 本文采用“一对多”策略. 设经训练后每个分类器的分类函数为 $f_j(x)$, $j=1, 2, \dots, n$. 对于任意待分类场景的特征矢量 x , 存在 $k \in \{1, 2, \dots, n\}$, 使得 $f(x) = \max_{j=1, \dots, n} f_j(x)$, 则判定该场景属于第 k 类.

训练 SVM 分类器常用的方法有 Chunking 算法^[14]、Decomposition 算法^[15]、SMO (序贯最小优化) 算法^[16]. 其中, 尤以 SMO 最为实用, 该方法将一个大型的 QP (Quadratic Programming) 问题分解为一系列仅含有两个 Lagrange 乘子 λ_1, λ_2 的最小子问题进行求解, 大大减少了计算时间和计算复杂度. 本文使用 SMO 算法进行分类器的训练.

4 实验结果 (Experiment results)

在我们自行设计研制的机器人 CASIA-I 上 (如图 1 所示) 对上述定位算法进行了实验验证.

图 1 CASIA-I 移动机器人

Fig. 1 CASIA-I mobile robot

利用装在 CASIA-I 上的具有仰俯和旋转功能的单目 CCD 对实验室、门厅和走廊 3 个场景, 在不同时间不同光照度条件下采集了 80 幅图像作为学习样本

训练 SVM,并且采集 15 幅图像作为测试样本.图 2 给出了 3 类场景的部分样本图像.

通过文中给出的滤波器组,对每个训练样本提取 40 个纹理特征,表 1 给出 3 类场景训练样本特征

矢量的均值和方差.对特征矢量进行 SVM 学习后,确定出不同场景之间的最优分类面以及相应的支持向量.本文对 3 种常用的核函数分别进行了实验,表 2 给出了实验结果.

图 2 室内、走廊、门厅 3 类场景的部分样本图像

Fig. 2 Images' samples of indoor, corridor and doorway

表 1 3 类场景训练样本特征矢量 (均值 E 与方差 D)

Table 1 Means and variances of textural feature vectors of three classes' training images (E : mean and D : variance)

参数	$f=2 \quad \theta=0$		$f=2 \quad \theta=\pi/4$		$f=2 \quad \theta=\pi/2$		$f=2 \quad \theta=3\pi/4$		$f=4 \quad \theta=0$	
	E	D	E	D	E	D	E	D	E	D
实验室	29.2512	27.9491	12.5411	11.3545	37.0791	34.5303	12.9598	12.3043	12.5589	15.4619
门厅	34.3346	26.6248	12.5524	11.3201	40.0447	31.8396	14.0246	12.8965	18.1464	17.2547
走廊	36.7383	27.2314	14.3634	13.2547	33.9532	30.6582	12.2316	11.7744	14.8181	12.4844
参数	$f=4 \quad \theta=\pi/4$		$f=4 \quad \theta=\pi/2$		$f=4 \quad \theta=3\pi/4$		$f=8 \quad \theta=0$		$f=8 \quad \theta=\pi/4$	
	E	D	E	D	E	D	E	D	E	D
实验室	5.4629	5.9058	18.4716	24.0518	5.7013	6.6566	5.8010	8.1502	2.6355	3.2363
门厅	5.4739	5.9671	19.9650	23.6465	5.3774	5.5778	7.9469	8.9444	2.4463	2.9350
走廊	5.8181	6.6665	16.8581	23.0588	5.0112	5.9016	8.6148	9.3183	2.5592	3.2128
参数	$f=8 \quad \theta=\pi/2$		$f=8 \quad \theta=3\pi/4$		$f=16 \quad \theta=0$		$f=16 \quad \theta=\pi/4$		$f=16 \quad \theta=\pi/2$	
	E	D	E	D	E	D	E	D	E	D
实验室	8.5973	14.2843	2.6261	3.4523	2.8645	4.5417	1.3878	1.7395	3.1218	4.9549
门厅	9.4808	13.9285	2.4973	2.9891	3.9055	5.3478	1.2237	1.5051	3.8620	6.1261
走廊	7.5495	14.2054	2.2283	2.8104	4.4039	5.8261	1.4406	1.9274	2.6421	4.8883
参数	$f=16 \quad \theta=3\pi/4$		$f=32 \quad \theta=0$		$f=32 \quad \theta=\pi/4$		$f=32 \quad \theta=\pi/2$		$f=34 \quad \theta=3\pi/4$	
	E	D	E	D	E	D	E	D	E	D
实验室	1.4120	1.8517	1.6056	2.8844	0.9273	1.1225	2.1684	4.6024	0.9326	1.1424
门厅	1.2532	1.5809	2.0485	3.4797	0.8211	0.9628	2.4913	5.0084	0.8194	0.9576
走廊	1.1836	1.3776	2.5105	4.1145	0.9230	1.1648	1.8778	4.4457	0.8945	1.1111

表 2 3 种核函数取不同参数的识别率

Table 2 Recognition rates of three kernels with different parameters

测试样本	识别率 (%)								
	多项式核函数			RBF核函数			Sigmoid核函数		
	$d=2$	$d=3$	$d=4$	$\delta=0.01$	$\delta=0.001$	$\delta=0.005$	$v=2 \quad c=0.9$	$v=2 \quad c=1$	$v=2 \quad c=1.1$
实验室	86.67	80.00	80.00	86.67	93.33	93.33	80.00	86.67	80.00
门厅	80.00	73.33	80.00	80.00	80.00	80.00	73.33	80.00	73.33
走廊	93.33	93.33	93.33	80.00	100	93.33	86.67	93.33	80.00
平均	86.67	82.22	84.44	82.22	91.11	88.89	80.00	86.67	77.78

由表 2 可知,取多项式核函数,参数 $d=2$ 时,3 类场景的识别率最高,平均识别率为 86.67%;取径向基函数,参数 $\delta=0.001$ 时,3 类场景的识别率最高,平均为 91.11%;取 Sigmoid 核函数,参数 $v=2, c=1$ 时,3 类场景的识别率最高,平均为 86.67%。比较这 3 种核函数可知,RBF 核函数要优于多项式核函数和 Sigmoid 核函数。鉴于此,本文选择 RBF 核函数的 SVM 分类器进行移动机器人的场景分类识别,获得了较高的定位准确率,并且满足机器人实时定位的要求。

5 结论 (Conclusion)

本文提出了一种基于全局特征场景识别的移动机器人定位方法,通过 Gabor 滤波器提取图像的全局纹理特征,并使用 SVM 对其进行分类,实现基于图像的场景识别,由于每类图像对应于机器人工作环境中的某个位置,从而实现了机器人的逻辑定位。实验验证:使用该算法机器人能达到 91.11% 正确定位率,并且能够满足移动机器人实时定位的要求。

参考文献 (References)

- [1] Ulrich I, Nourbakhsh I. Appearance-based place recognition for topological localization [A]. Proceedings of the IEEE International Conference on Robotics and Automation[C]. 2000. 1023 - 1029.
- [2] Zhou C, Wei Y C, Tan T N. Mobile robot self-localization based on global visual appearance feature [A]. Proceedings of the 2003 IEEE International Conference on Robotics and Automation[C]. 2003. 1271 - 1276.
- [3] Carneira M J, Orwell J, Tumes R, et al. Perceptual grouping from Gabor filter responses [A]. Proceedings of the Ninth British Machine Vision Conference[C]. Southampton, UK: 1998. 336 - 345.
- [4] Manjunath B, Ma W Y. Textures features for browsing and retrieval of image data [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837 - 842.
- [5] Lee T S. Image representation using 2D gabor wavelets [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(10): 959 - 971.
- [6] Rui Y, Huang T S, Chang S F. Image retrieval: past, present, and future [J]. Journal of Visual Communication and Image Representation, 1999, 10(1): 1 - 23.
- [7] Vapnik V N. Statistical Learning Theory [M]. Wiley, New York: 1998.
- [8] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Methods [M]. Cambridge, UK: Cambridge University Press, 2000.
- [9] Bovik A C, Clark M, Geisler W B. Multichannel texture analysis using localized spatial filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(1): 55 - 73.
- [10] Tan T N. Texture feature extraction via cortical channel modeling [A]. Proceedings of the 11th IAPR International Conference on Pattern Recognition[C]. 1992, 3. 607 - 610.
- [11] Motton L, Cortes C, et al. Comparison of classifier methods: a case study in handwriting digit recognition [A]. Proceedings of the IEEE International Conference on Pattern Recognition[C]. 1994. 77 - 87.
- [12] K̄r̄bel U. Pairwise Classification and Support Vector Machines [M]. Cambridge, MA: MIT Press, 1999. 255 - 266.
- [13] Hsu C W, Lin C J. A comparison of methods for multi-class support vector machines [J]. IEEE Transactions on Neural Networks, 2002, 13(2): 415 - 425.
- [14] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers [A]. Proceedings of the Fifth Annual Workshop on Computational Learning Theory[C]. Pittsburgh, PA, USA: 1992. 144 - 152.
- [15] Osuna E, Freund R, Girosi F. An improved training algorithm for support vector machines [A]. Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing[C]. New York: IEEE Press, 1997. 276 - 285.
- [16] Platt J C. Fast training of Support Vector Machines Using Sequential Minimal Optimization [M]. Cambridge, MA: MIT Press, 1999. 185 - 208.

作者简介:

李桂芝 (1971-), 女, 博士研究生. 研究领域: 机器人视觉, 多传感器数据融合.

安成万 (1974-), 男, 博士研究生. 研究领域: 图像处理, 机器人视觉, 多传感器数据融合.

杨国胜 (1963-), 男, 副教授, 博士后. 研究领域: 多传感器多目标数据融合技术, 复杂系统与智能控制等.