

SVM 用于基于内容的自然图像分类和检索

付 岩¹⁾ 王耀威²⁾ 王伟强¹⁾ 高 文^{1),2)}

¹⁾(中国科学院计算技术研究所 北京 100080)

²⁾(中国科学院研究生院 北京 100039)

摘 要 在传统的基于内容图像检索的方法中,由于图像的领域较宽,图像的低级视觉特征和高级概念之间存在着较大的语义间隔,导致检索效果不佳.该文认为更有现实意义的做法是,缩窄图像的领域以减小低级特征和高级概念间的语义间隔,并利用机器学习方法自动建立图像类的模型,从而提供用户概念化的图像查询方式.该文以自然图像领域为例,使用支持向量机(SVM)学习自然图像的分类,学习到的模型用于自然图像分类和检索.实验结果表明作者的方法是可行的.

关键词 支持向量机;基于内容的图像检索;图像分类;特征不变性

中图法分类号 TP391

Content-Based Natural Image Classification and Retrieval Using SVM

FU Yan¹⁾ WANG Yao-Wei²⁾ WANG Wei-Qiang¹⁾ GAO Wen^{1),2)}

¹⁾(*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080*)

²⁾(*Graduated School of the Chinese Academy of Sciences, Beijing 100039*)

Abstract In the traditional approach of content-based image retrieval, the wide image domain results in the wide semantic gap between the low-level features and the high-level concepts. We propose to narrow the image domain and use machine learning methods to automatically construct models for image classes, thus providing users with a conceptualized way to image query. In this paper, support vector machines are trained for natural image classification. The resulting image class models are incorporated into image retrieval system, so that the users can search natural images by classes. The experimental results are promising.

Keywords support vector machines; content-based image retrieval; image classification; feature invariance

1 引 言

由于数字图像数量的猛增,对其进行快速、高效检索的要求愈加强烈.从 20 世纪 90 年代初开始,基于内容的图像检索技术得到了广泛发展.本文介绍的方法的目的在于对自然图像的内容进行学习,并用于分类和检索.

基于内容的图像检索从如何赋予视觉特征以语

义的角度可以分为两类^[1]:一类是基于相似度的方法.通常由用户指定一个样例图像,然后比较这个样例图像与数据库中的图像的视觉特征,相近的作为查询结果.这类系统包括 QBIC^[2],Virage^[3],VisualSeek^[4]等等.这类系统一般是通用的查询系统,对图像的领域没有限制.但是,具有相同或相似低级视觉特征的图像可能有着完全不同的概念信息.也就是说,在图像的低级特征和高级概念之间存在着语义间隔(semantic gap)^[1].图像的领域越宽,间隔就

收稿日期:2002-04-08;修改稿收到日期:2003-03-18.本课题得到国家“八六三”高技术研究发展计划项目“实时图像检索与过滤关键技术研究”课题(2001AA142140)和国家网络与信息安全中心“网络视频流内容检测技术研究及系统”课题资助.付 岩,男,1977 年生,博士研究生,主要研究方向为模式识别、机器学习. E-mail: yfu@ict. ac. cn. 王耀威,男,1976 年生,博士研究生,主要研究方向为多媒体技术、模式识别. 王伟强,男,1972 年生,博士,副研究员,主要研究方向为多媒体技术、模式识别. 高 文,男,1956 年生,博士,教授,博士生导师,主要研究领域为人工智能应用和多媒体技术,侧重于计算机视觉、模式识别与图像处理、多媒体数据压缩、多模式接口以及虚拟现实等的研究.

越大. 所以这类系统虽然取得了一定成功, 但检索效果却常常不如人意.

另一类方法是基于解释的方法. 即在分析图像的内容并提取特征之后, 并不进行比较, 而是直接给出语义解释, 使得用户可以进行概念化的查询, 如文献[5,6]中的方法便属于该类. 文献[5]对图像集采取了多级分类的方法, 比如首先将图像分为室内和室外两类, 再把室外图像分为城市和风景两类, 风景再进一步分为落日、山脉、森林等等. 这样就降低了分类的难度, 要么在宽图像领域中进行粗分类, 要么在窄图像领域中进行细分类. 在每一级分类时, 采取了贝叶斯分类的方法. 它假设在每一级分类时, 图像集包含的图像类别是固定的; 又假设每类图像的先验概率是已知的, 但实际上这两个理论假设在实际应用中是难以满足的. 文献[6]提出了基于布局(configuration)的自然图像分类方法. 它把图像中的各个区域之间定性的空间关系作为特征, 从而在自然图像范围内, 能够捕获图像的概念信息. 比如, 如果图像的颜色分布为上部蓝色、中部白色、下部黄褐色, 那么这样一种布局暗示着图像可能是雪山类的景物. 这种方法的缺点在于, 它完全依赖于人对图像类别的先验知识进行人工建模, 而且限于某一图像领域. 当图像领域很宽时, 需要人工建模的图像类别的数量将是不可接受的.

我们认为概念化的查询更有现实意义. 所以, 限制图像的领域, 缩小低级特征和高级概念之间存在的语义间隔, 并借助机器学习的方法自动建立图像类的模型用于图像检索, 可能是一条有意义的道路. 我们的方法就是采用支持向量机(SVM)作为学习机器, 在自然图像领域中学习图像的内容, 学习到的图像类别模型用于图像检索. 与文献[5]相比, 我们的方法并不要求图像类别是固定的, 也不要求图像类的先验概率是已知的, 从而更有实际意义. 与文献[6]的不同之处在于, 在建立图像类的模型时我们回避了对显式先验知识的需求, 学习到的模型被隐含在了学习机器之中(这里是支持向量). 另外, 我们也对特征的选取做了实验和比较.

本文第2节给出描述图像所使用的视觉特征; 第3节介绍我们采用的学习算法; 第4节是实验的设计和结果; 第5节给出几点总结以及进一步的工作.

2 视觉特征选取

由于颜色是自然景物图像中最重要的内容信息, 我们选取了颜色直方图(color histogram)、颜色

一致向量(color coherence vector)和颜色相关图(color correlogram)作为图像的特征. 这3种特征依次具有不同的区分能力和不变性等性质, 我们将在2.6节中作以比较.

2.1 颜色空间

颜色空间是颜色的表示方法, 可以分为面向硬件的和面向用户的两种^[7]. 面向硬件的颜色空间包括RGB, CMY, YIQ等. 其中RGB是现在数字图像广泛采用的颜色空间. 这些颜色空间适合于硬件显示, 但并不直接与色调、饱和度和亮度等人的主观感觉相对应. 面向用户的颜色空间包括HIS, HCV, HSV, HSB以及MTM, $L^*u^*v^*$, $L^*a^*b^*$ 和 $L^*C^*H^*$ 等. 这些颜色空间与人的感觉相符合. 我们采用了HSV颜色空间. 它由3个分量组成, 分别代表色调H(hue)、饱和度S(saturation)、亮度V(value). 我们采用了如下的从RGB到HSV的变换方法:

$$\left\{ \begin{array}{l} Max = \max(R, G, B) \\ Min = \min(R, G, B) \\ V = 0.299R + 0.587G + 0.114B \\ S = \begin{cases} 0, & \text{如果 } Max = 0 \\ (Max - Min)/Max, & \text{否则} \end{cases} \\ H = \begin{cases} 0, & \text{如果 } Max = Min \\ 60(G - B)/(Max - Min), & \text{如果 } Max = R \text{ 并且 } G > B \\ 360 + 60(G - B)/(Max - Min), & \text{如果 } Max = R \text{ 并且 } G < B \\ 60(2 + (B - R)/(Max - Min)), & \text{如果 } Max = G \\ 60(4 + (R - G)/(Max - Min)), & \text{否则} \end{cases} \end{array} \right. \quad (1)$$

其中, $R, G, B, S, V \in [0, 1]$, $H \in [0, 360]$. 我们将这个变换表示为

$$T: RGB \rightarrow HSV \quad (2)$$

这里, 我们采用的变换方法与MPEG-7标准中的变换方法^[8]的不同在于亮度V的计算方法. 在MPEG-7中, V取R, G, B中的最大值, 而我们认为我们所采用的方法更有合理性, 因为它考虑了不同RGB分量对亮度的贡献程度.

2.2 颜色量化

由于这三种特征都需要对颜色出现的频率进行统计, 所以要对颜色空间量化. 这里, 我们将HSV颜色空间量化成256个颜色. 方法是H分量量化成16个值, S和V分量分别量化成4个值. 我们将这种量化表示为

$$Q: HSV \rightarrow C \quad (3)$$

其中 $C = \{c_i | i = 0, 1, 2, \dots, 255\}$, c_i 表示第 i 种量化颜色.

2.3 颜色直方图

颜色直方图^[9]由 Swain 和 Ballard 提出,并最早用于基于颜色的图像检索. 设原始图像用 $I(x, y)$ 表示,则颜色直方图 H 可以定义如下:

$$H = \{h_0, h_1, \dots, h_{255}\} \quad (4)$$

$$h_k = \frac{\sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} \begin{cases} 1, & \text{如果 } Q(T(I(x, y))) = c_k \\ 0, & \text{否则} \end{cases}}{N_1 \times N_2} \quad (5)$$

其中 N_1 和 N_2 表示图像的宽和高, $k = 0, 1, \dots, 255$.

直观地说, h_k 是图像中具有量化颜色 c_k 的像素数量的比例.

2.4 颜色一致向量

颜色直方图简单地对颜色出现的频率进行统计,没有考虑颜色的空间分布情况. 因而它的区分能力很弱,不能区分颜色密集分布和松散分布的情况. 比如满地的落叶和日落的景象可能具有类似的颜色直方图,但颜色分布却极为不同. 颜色一致向量^[10]通过考虑相同颜色的像素的连通程度,引入了一定的颜色空间分布信息,而更具有区分能力. 它实际上是将颜色直方图中的每个 h_k 分成两个值,一个表示连通度高的像素,另一个表示连通度低的像素. 颜色一致向量 S 可以定义如下:

$$S = \{(s_0, s'_0), (s_1, s'_1), \dots, (s_{255}, s'_{255})\} \quad (6)$$

$$s_k = \frac{\sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} \begin{cases} 1, & \text{如果 } Q(T(I(x, y))) = c_k, \text{ 且像素} \\ & (x, y) \text{ 所在同色区域大于图像的 } 1\% \\ 0, & \text{否则} \end{cases}}{N_1 \times N_2} \quad (7)$$

s'_k 由类似方法定义. 从定义可知有 $h_k = s_k + s'_k, k = 0, 1, \dots, 255$.

直观地说, s_k 是图像中具有量化颜色 c_k 并具有高连通度的像素数量的比例.

2.5 颜色相关图

与颜色直方图和颜色一致向量不同的是,颜色相关图^[11]不仅从颜色出现的频率角度描述图像的颜色分布信息,而且刻画了图像中各种颜色之间的空间分布关系. 它是通过统计图像中两种颜色在各种距离上的变化关系来实现的. 颜色相关图可以有多种形式,考虑到实际可以接受的空间和时间复杂度,我们采用一种简化的常用形式,即颜色自相关图 A . 它定义如下:

$$A = \{a_i^k | i = 0, 1, \dots, 255; k = 1, 2, 3, 4, 5\} \quad (8)$$

$$a_i^k = \frac{\Pr_{Q(T(I(x_1, y_1))) = c_i, (x_2, y_2) \in I} [Q(T(I(x_2, y_2))) = c_i]}{1} \quad (9)$$

$$= c_i | | (x_1, y_1) - (x_2, y_2) | = k] \quad (9)$$

其中 $| (x_1, y_1) - (x_2, y_2) |$ 是像素 (x_1, y_1) 和 (x_2, y_2) 之间的距离.

直观地说, a_i^k 就是与具有量化颜色 c_k 的像素距离为 k 的像素仍具有量化颜色 c_k 的概率或比例.

2.6 特征不变性

每个图像类中的图像都包含使之从属于该类的内容,同时又包含与该类无关的信息,这些内容是千变万化的. 好的特征应该能够排除噪音,抓住图像类的本质. 特征的这种性质可以用不变性来刻画. 不变性^[12]可以简单地理解为特征在描述图像的内容时不受图像某类形变的影响的性质. 一般来讲,对于低级特征,包含的信息越多,区分能力越强,不变性越低. 选择特征的准则就是在一定的图像领域下,选择具有适当不变性的特征,使其能够把同类图像归为一类,把不同类图像区分开来. 表 1 从颜色分布的角度对上述 3 种特征的不变性、区分能力等性质作了比较.

表 1 3 种特征的比较

	包含的信息	区分能力	不变性	计算复杂度
颜色直方图	少	低	强	低
颜色一致向量	中	中	中	中
颜色相关图	多	强	低	高

3 学习算法

我们选择 SVM^[13] 作为学习机器,是因为 SVM 不仅有统计学习理论^[14] 作为其推广性的保证,而且在很多实际问题中获得了优于其他方法的结果^[15]. 下面简要对 SVM 做简要介绍.

3.1 线性 SVM

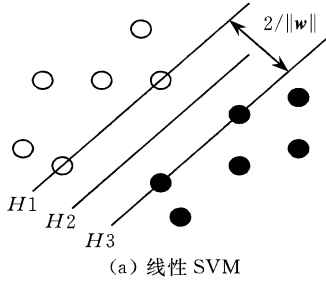
在学习样本是线性可分的情况下,为简单起见,考虑二维的情况. 在图 1(a) 中,实心点和空心点代表两类样本, H 为分类线, $H1, H2$ 分别为过各类中离分类线最近的样本且平行于分类线的直线,它们之间的距离叫做分类间隔. 所谓最优分类线就是要求分类线不但能将两类正确分开,而且使分类间隔最大. 分类线方程为 $x \cdot w + b = 0$, 我们可以对它进行归一化,使得线性可分的样本集 $(x_i, y_i), i = 1, 2, \dots, n, x \in R^2, y \in \{+1, -1\}$, 满足

$$y_i [(w \cdot x_i) + b] - 1 \geq 0, i = 1, 2, \dots, n \quad (10)$$

此时分类间隔等于 $2/\|w\|$, 使间隔最大等价于使 $\|w\|$ 最小. 满足条件(10)且使 $\|w\|$ 最小的分类面就叫做最优分类面, $H1, H2$ 上的训练样本点,也就是使式(10)中等号成立的样本点称作支持向量. 解这个最优化问题便得到最优分类函数:

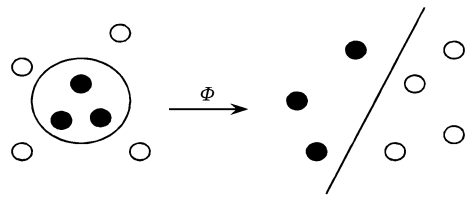
$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right\} \quad (11)$$

使分类间隔最大实际上就是在保持经验风险固定(等于0)的前提下,降低学习机器的 VC 维,控制



(a) 线性 SVM

学习机器的复杂性,以得到好的推广能力,这是 SVM 的核心思想之一。



(b) 非线性 SVM

图 1

3.2 非线性 SVM 与核函数

在学习样本是线性不可分,但却是非线性可分的情况下,我们可以通过非线性变换把学习样本变换到高维空间,使其在高维空间里是线性可分的(图 1 (b)).但这会导致特征空间的维数骤然增加,也就是“维数灾难”问题.注意到上面最优化问题和最优分类函数中只涉及样本特征向量的点积,那么就可以利用核函数技术避免实际去做这样的非线性变换.用核函数 $K(x, y)$ 代替原来的点积 $(x \cdot y)$. Mercer 定理指出,核函数 $K(x, y)$ 通过与其相联系的非线性变换 Φ 隐含地把特征向量映射到高维特征空间,使得学习样本成为线性可分的.常用的核函数包括多项式核函数 $K(x, y) = [(x \cdot y) + s]^d$, 径向基函数 $K(x, y) = \exp(-\sigma \|x - y\|^2)$, Sigmoid 函数 $K(x, y) = \tanh(k(x \cdot y) - \mu)$ 等等.选定一个核函数后,最优分类函数成为 $f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i K(x_i, x) + b^* \right\}$.

4 实 验

依据式(1)~(9)计算图像的 3 种颜色特征向量,其中颜色直方图为 256 维,颜色一致向量为 512 维,颜色自相关图为 1280 维.这 3 种颜色特征分别作为支持向量机的输入向量,对图像类进行学习.我们选取了落日、原野和海洋 3 个图像类.这是个多类分类问题,我们这里采用 one vs. rest 的方法对每种图像类别进行学习.实验数据选自 Corel 图像库,每类图像有正例 150 幅,以 8:7 的比例分配给训练集和测试集.不同类别的图像互相作为反例加入训练集中.训练集的反例还包括 500 幅除此 3 类图像以外的其他类图像.测试集一共 800 幅,其中除了 3 种待分类图像外,还包括其他类图像.在进行训练时,分别尝试上一节中提到的 3 种核函数,并最终选择具有最好实验效果的核函数.

训练后的 SVM 被包含进图像检索系统中,使

得用户可以进行基于图像类别的自然图像检索.表 2 给出了实验结果.

表 2 在自然图像领域内,采用不同特征和 SVM 的实验结果
(a) 用颜色直方图作为特征的检索结果

图像类别	SVM 类型	准确率 (%)	查重率 (%)
Sunset	L	66.35	98.57
	R ($\sigma=5$)	68.00	97.14
Field	L	40.48	38.64
	R ($\sigma=4$)	78.38	65.91
Sea	L	71.08	84.29
	R ($\sigma=2$)	72.29	85.71
Average	L	59.30	73.83
	K	72.89	82.92

(b) 用颜色一致向量作为特征的检索结果

图像类别	SVM 类型	准确率 (%)	查重率 (%)
Sunset	L	78.41	98.57
	R ($\sigma=2$)	81.40	100
Field	L	46.67	31.82
	R ($\sigma=6$)	64.52	43.18
Sea	L	83.58	80.00
	R ($\sigma=2$)	83.33	78.57
Average	L	69.55	70.13
	K	76.42	73.92

(c) 用颜色相关图作为特征的检索结果

图像类别	SVM 类型	准确率 (%)	查重率 (%)
Sunset	L	98.59	100
	P ($s=1, d=2$)	100	100
Field	L	46.88	68.18
	P ($s=1, d=2$)	54.10	75.00
Sea	L	74.65	75.71
	P ($s=1, d=2$)	75.00	73.86
Average	L	73.37	81.30
	K	76.36	82.95

其中 L 代表线性, K 代表非线性, R 代表径向基函数, P 代表多项式函数, 后面跟随的是选取的参数

从表 2 中可以看出,对于落日图像类,在采用颜色相关图时,得到了出人意料的极佳检索结果,而原

野图像类在采用颜色直方图时得到了较好结果. 海洋则是在采用颜色一致向量时得到了较好结果. 这样的实验结果在表面上似乎并不令人乐观, 因为难以从中看出什么规律出来. 然而, 我们认为这是很重要的一个结果. 这说明, 一方面我们不能采用唯一的特征来描述所有自然图像类; 另一方面, 也是很重要的一点, 那就是分类的效果并不一定随着特征的复杂程度的上升而变得更好. 这正是不变性概念的意义所在.

另外, 非线性 SVM 在几乎所有分类中都比线性 SVM 得到了更好的分类结果. 虽然我们只用了很少的训练数据, 但仍然获得了令人满意的效果. 平均来看, 适当选取图像特征和核函数, 可以得到比较满意的查准率和查全率.

5 结论和未来的工作

从实验的结果, 我们可以得出结论, 用 SVM 作为学习机器学习自然图像的内容并分类的方法是可行的, 学习到的模型可以用于基于概念的图像检索. 有两点需要指出, 第一, 图像特征的选取, 要依据不变性原则, 特征包含的信息过多或过少都会影响正确分类. 第二, 缩小图像的领域范围, 通过学习自动建立图像类的模型, 可能是在一定程度上解决基于内容的图像内容检索中低级特征与高级概念之间的语义间隔的有效途径.

当然, 我们的做法也有不足, 这里只选用了图像的全局特征, 只在少量图像类上作了实验. 同时, 宽领域下的语义间隔问题本身可能是无法彻底解决的. 所以, 为了使我们的方法能够用于更广的范围和得到更好的效果, 还需要进一步的工作. 这包括引入更多的图像信息(如纹理、形状等)以及图像的分割等等. 另外, 把对图像的先验知识融入适当构造的核函数中, 以提高分类的效果也是进一步的研究内容.

参 考 文 献

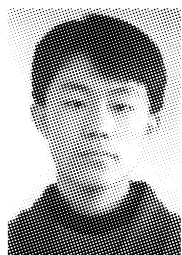
1 Smeulders A, Worring M *et al.* Content-based image retrieval

- at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(12):1349~1380
- 2 Flickner M *et al.* Query by image and video content: The QBIC system. *IEEE Computer*, 1995, 28(9):23~32
- 3 Bach J R, Fuller C, Gupta A *et al.* Virage image search engine: an open framework for image management. *SPIE Storage and Retrieval of Image and Video DataBases*, 1996, 4:76~87
- 4 Smith J, Chang S F. VisualSEEK: A fully automated content-based image query system. In: *Proceedings of the 4th ACM Multimedia Conference*, Boston MA, USA, 1996. 87~98
- 5 Vailaya A, Figueiredo M, Jain A, Zhang H-J. A Bayesian framework for semantic classification of outdoor vacation images. In: *Proceedings of SPIE: Storage and Retrieval for Image and Video Databases VII*, San Jose, CA, USA, 1999, 3656: 415~426
- 6 Lipson P, Grimson E, Sinha P. Configuration based scene classification and image indexing. In: *Proceedings of the 16th IEEE Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997. 1007~1013
- 7 Alberto B. *Visual Information Retrieval*. San Francisco, California: Morgarn Kaufmann Publishers, 1999. 81~81
- 8 Text of ISO/IEC 15938-3/FCD Information Technology—Multimedia Content Description Interface—Part 3 Visual. Singapore, 2001
- 9 Swain M J, Ballard D H. Color indexing. *International Journal of Computer Vision*, 1991, 7(1): 11~32
- 10 Pass G, Zabih R. Histogram refinement for content-based image retrieval. In: *Proceedings of the IEEE Workshop on Application of Computer Vision*, Sarasota, Florida, 1996. 96~102
- 11 Huang J *et al.* Image indexing using color correlogram. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Puerto Rico, 1997. 762~768
- 12 Burkhart H, Siggelkow S. Invariant features for discriminating between equivalence classes. In: *Nonlinear Model-based Image Video Processing and Analysis*. NY: John Wiley and Sons, 2000
- 13 Schölkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, Mass: MIT Press, 2002
- 14 Vapnik V N. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 2000
- 15 Schölkopf B, Burges C J C, Smola A J. *Advances in Kernel Methods—Support Vector Learning*. Cambridge, MA: MIT Press, 1999

recognition.

WANG Wei-Qiang, born in 1972, Ph. D., associate professor. His research interests include multimedia techniques and pattern recognition.

GAO Wen, born in 1956, Ph. D., professor and Ph. D. supervisor. His research interests include multimedia data compression, image processing, computer vision, multimodal interface, artificial intelligence and virtual reality.



FU Yan, born in 1977, Ph. D. candidate. His research interests include pattern recognition and machine learning.

WU Yao-Wei, born in 1976, Ph. D. candidate. His research interests include multimedia techniques and pattern