

# 基于支持向量机的语义图像分类研究

李晶, 姚明海

(浙江工业大学 信息工程学院, 浙江 杭州 310014)

**摘要:**随着多媒体数据库的不断发展,传统的利用关键词进行图像检索已经越来越不能满足图像检索的需要,基于内容的图像检索已成为当前的研究热点。对图像的颜色和纹理特征进行提取,融合图像的颜色和纹理特征作为图像的特征向量,用支持向量机实现图像的低层特征和高级语义间的关联。实验结果表明,多特征的图像检索要比单一的特征检索效果好,在HSV颜色特征的基础上引入灰度共生矩阵纹理特征后可有效提高检索效率,而且采用支持向量机融合多特征可成功用于图像的语义的检索。

**关键词:**支持向量机;HSV颜色特征;灰度纹理特征;灰度共生矩阵

**中图分类号:**TP311

**文献标识码:**A

**文章编号:**1673-629X(2010)02-0075-04

## Research of Semantic Image Classification Based on Support Vector Machine

LI Jing, YAO Ming-hai

(Information Engineering College, Zhejiang University of Technology, Hangzhou 310014, China)

**Abstract:** With the continually development of multimedia database, traditional image retrieval method which is based on keywords can not satisfy most of the requirement of image retrieval. In recent years, more and more researchers have transferred their focus of scientific researches on content-based image retrieval. Extract the color and texture feature of the image, integrate color feature and texture feature as the feature vectors. Use support vector machine correlating image low-level feature with high-level semantic. Experimental result shows that, the retrieval of image with multiple features is better than that of image with single feature. The introduction of gray scale texture feature into HSV color feature can improve the efficiency of image retrieval. Furthermore, support vector machine can be successfully used in image semantic retrieval.

**Key words:** support vector machine; HSV color feature; gray scale texture feature; GLCM

### 0 引言

基于内容的图像检索技术<sup>[1]</sup>由于具有传统的文本检索技术所无法比拟的优越性,近年来,已经成为了一个非常活跃的研究领域。从目前的研究现状来看,基于内容的图像检索可分为3层。下层是基于颜色、纹理、形状等反映图像基本物理特征的检索,是最直接也是最基本的层面,用到图像信息处理、图像分析和相似性匹配技术<sup>[2]</sup>。中间层是基于图像对象语义的检索,如图像中实体及实体之间的空间关系的检索,对象级检索技术建立在下层特征基础上,并引入了对象模型库、对象识别和人工智能等图像理解技术<sup>[3]</sup>。最上层是基于图像概念级语义的检索,其技术建立在对象层

语义特征提取的基础上,引入了对象和场景之间的逻辑、情感等高层语义的描述及识别,需要用到知识库和更加有效的人工智能和神经网络技术<sup>[4]</sup>。

目前,基于内容的图像检索主要是利用图像的颜色、纹理、形状等低层的视觉特征达到检索的目的,其中利用颜色和纹理这两种特征进行检索应用较广泛,且检索的性能很好。

### 1 颜色特征提取

RGB颜色空间是由红、绿、蓝三种颜色按照一定的比例混合在一起,是加色法模型,它被广泛地使用。但它特别依赖设备。HSV颜色空间用H(亮度),S(饱和度),V(强度)三种元素表示每个像素的颜色组合,从而也可用三组二进制代码的集合(HSV色度空间值)表示一个像素。它可以独立感知各种颜色分量的变化,适合颜色的相似性比较。颜色的量化:如果直接

收稿日期:2009-05-28;修回日期:2009-08-30

作者简介:李晶(1985-),男,安徽宣城人,硕士研究生,研究方向为语义图像检索;姚明海,博士,教授,研究方向为企业集成信息化和智能信息处理。

在一副图像上提取颜色特征,计算量非常巨大,所以需要主要色彩表示。因此,进行颜色量化是统计颜色特征的必要操作。颜色量化的目的是减少图像中颜色的数码,使用尽可能少的颜色表达图像的颜色视觉信息。

### 1.1 HSV 变换

设  $uc$  是 RGB 颜色空间中的一个颜色点, $uc$  是 HSV 颜色空间中的一个颜色点,从 RGB 到 HSV 的变换  $T_c$  使得  $uc = T_c(uc)$ ,则变换  $T_c$  可以定义如下:

RGB $\Rightarrow$ HSV 对  $r, g, b \in [0, 1]$ ,  $T_c$  将根据下面的等式计算对应的  $h, s, v \in [0, 1]$ :

$$v = \max(r, g, b)$$

$$s = \frac{v - \min(r, g, b)}{v}$$

$$\text{定义 } r' = \frac{v - r}{v - \min(r, g, b)}$$

$$g' = \frac{v - g}{v - \min(r, g, b)}$$

$$b' = \frac{v - b}{v - \min(r, g, b)}$$

$$h' =$$

$$\left\{ \begin{array}{l} 5 + b' \text{ 若 } r = \max(r, g, b) \text{ 且 } g = \min(r, g, b) \\ 1 - g' \text{ 若 } r = \max(r, g, b) \text{ 且 } g \neq \min(r, g, b) \\ 1 + r' \text{ 若 } g = \max(r, g, b) \text{ 且 } b = \min(r, g, b) \\ 3 - b' \text{ 若 } g = \max(r, g, b) \text{ 且 } b \neq \min(r, g, b) \\ 3 + g' \text{ 若 } b = \max(r, g, b) \text{ 且 } r = \min(r, g, b) \\ 5 - r' \text{ 其它} \end{array} \right.$$

$$h = h' / 6$$

### 1.2 颜色的量化

如果直接在一副图像上提取颜色特征,计算量非常巨大,所以需要主要色彩来表示。因此,进行颜色量化<sup>[5]</sup>是统计颜色特征的必要操作。颜色量化的目的是养活图像中的颜色视觉信息。常用的有非间隔量化法和阶层聚类法。

阶层聚类方法是一种非监督学习方法,它序列地把  $n$  个样本分割到  $c$  个聚类中。分割的第一步是把每一个样本数据归并到一个聚类中;第二步是把  $n$  个聚类合并为  $n - 1$  个聚类,再把  $n - 1$  个聚类合并为  $n - 2$  个聚类,直到得到  $c$  个聚类。

文中使用非间隔量化法,把  $H, S, V$  三个分量按照人的颜色感知把色度  $H$  空间分成 8 份,饱和度  $S$  和亮度  $V$  空间分成三份,并根据色彩的不同范围进行量化,把量化后的 3 个颜色分量合成一维特征矢量:  $G = 9H + 3S + V$ ,这样可以获得一个 72 维的特征向量。

## 2 纹理特征提取

纹理特征是图像的重要特征之一,其本质是刻画

像素的领域灰度空间分布规律。图像纹理分析是图像分析处理研究中的一个重要组成部分。纹理分析在场景分析、医学图像分析、遥感图像分析等领域有着重要应用。灰度共生矩阵(Gray Level Co-occurrence Matrix, GLCM)<sup>[6-8]</sup>是分析纹理特征的重要方法。GLCM 描述了图像中,在  $\theta$  方向上距离为  $d$  的一个分别具有灰度  $i$  和  $j$  的像素出现的概率。假定待研究的纹理区域是矩形的,其在水平方向有  $N_x$  个分辨率,在垂直方向上有  $N_y$  个分辨率,图像的灰度级为  $N, L_x = \{0, 1, 2, \dots, N_x - 1\}, L_y = \{0, 1, 2, \dots, N_y - 1\}$  分别为水平和垂直空间域。灰度  $i$  和  $j$  的一对像素点位置方向为  $\theta$ , 距离  $d$  的概率记为  $p_{i,j(d,\theta)}$ , 具体计算公式为:

$$p(i, j, d, 0^\circ) = \# \{ (k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid k - m = 0, \mid l - n \mid = d;$$

$$f(k, l) = I, f(m, n) = j \}$$

$$p(i, j, d, 45^\circ) = \# \{ (k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = d)$$

$$\text{或 } (k - m = -d, l - n = -d); f(k, l) = i,$$

$$f(m, n) = j \}$$

$$p(i, j, d, 90^\circ) = \# \{ (k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid k - m \mid = d, l - n = 0;$$

$$f(k, l) = I, f(m, n) = j \}$$

$$p(i, j, d, 135^\circ) = \# \{ (k, l), (m, n) \in (L_y \times L_x) \times (L_y \times L_x) \mid (k - m = d, l - n = -d)$$

$$\text{或 } (k - m = -d, l - n = -d); f(k, l) = i,$$

$$f(m, n) = j \}$$

记号  $\# \{x\}$  表示集合  $x$  的元素数。矩阵的第  $i$  行  $j$  列元素表示所有  $\theta$  方向,相邻间隔为  $d$  的像素中有一个取  $i$  值,另一个取  $j$  值的相邻对点数。

一幅图像的灰度共生矩阵反映了图像灰关于方向、相邻间隔、变化幅度的统合信息,它是分析图像局部模式和它们排列规则的基础。从它出发,可以进一步提取描述图像纹理的一系列特征。文中主要提取了纹理、熵、惯性矩和局部平稳性四个最有效的纹理特征组成特征向量。

基于灰度共生矩阵的纹理特征提取算法如下:

(1) 构造 4 个方向的灰度共生矩阵  $p(1, 0), p(0, 1), p(1, 1), p(1, -1)$ ;

(2) 计算共生矩阵的二阶矩、相关度、熵、对比度、逆差矩及和方差;

(3) 构造纹理特征向量。

## 3 SVM 的基本理论及关联方法

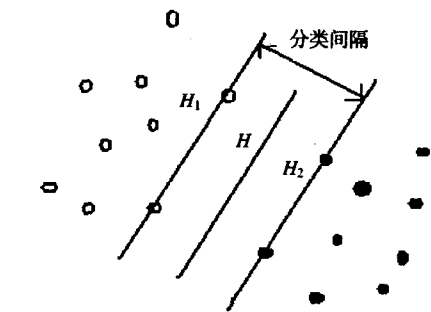
### 3.1 SVM 的基本理论

选择 SVM 作为学习机器,对图像检索中的底层

特征和高层语义之间进行关联。与文献[9]相比,文中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。支持向量机是在传统统计学基础上发展起来的一种具有坚实的机器学习方法,是目前针对小样本统计估计和预测学习的最佳理论。它遵循结构风险最小化原理,不需要特定问题的先验知识,在有限的训练样本情况下,可以很好地控制学习机器的推广能力。其算法实质上是求解一个凸规划问题或其对偶问题二次规划(QP)问题,对于最基本的两分类问题,存在线性可分和线性不可分两种 SVM,可简述为将输入空间中的样本通过某种非线性函数关系映射到一个特征空间中(维数可能较高),使两类样本(可推广到多类样本)在此特征空间中线性可分,并寻找样本在此特征空间中的最优分类超平面。

下面简要介绍一下 SVM 原理。

在学习样本是线性可分的前提下,可用两维情况说明。在图 1 中,实心点和空心点代表两类样本, $H$  为分类线, $H_1, H_2$  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。所谓最优分类就是要求分类线不但能将两类正确分开,而且合分类间隔最大。分类线方程为  $x^* w + b = 0$ , 可以对其进行归一化,使得线性可分的样本集  $(x_i, y_i), i = 1, 2, \dots, n, x \in R^d, y \in \{+1, -1\}$ , 满足  $y_i[(w * x_i) + b] - 1 \geq 0, i = 1, \dots, n$  (1)



线性可分情况下的最优分类线

图 1 SVM 原理图

此时分类间隔等于  $2 / \| w \|$ , 使间隔最大等价并使  $\| w \|^2$  最小。满足式(1)且使  $\| w \|^2$  最小的分类面就叫做最优分类面,  $H_1, H_2$  上的训练样本点, 也就是使式(1)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:

$$f(x) = \text{sgn}(\sum_{i=1}^n a_i y_i k(x_i, x) + b)$$

其中  $\text{sgn}()$  为符号函数, 通过参数的符号判断输入向量的类别;  $f(x)$  将给定的属于两个类别的训练样本分开, 构造平面的依据是两类样本离超平面的距离

最大化, 如图 1 所示。

用 SVM 实现图像分类<sup>[10]</sup>, 首先要从图像的抽取图像特征向量作为原始空间的输入向量, 由于图像的内容的非确定性和多样性, 即使提取的特征良好, 也不能保证图像线性可分, 即图像是非线性可分的。SVM 首先通过核函数将原始空间中低维数据映射到高维特征空间, 在高维特征空间构造最优分类平面, 即通过求解得到一个最优分类函数:

$$f(x) = \text{sgn}(\sum_{i=1}^n a_i y_i k(x_i, x) + b)$$

其中,  $k(x_i, x)$  为核函数, 其选取应使它为特征空间的一个点积, 即存在函数  $\Phi$ , 使  $\Phi(x_i) \times \Phi(x) = k(x_i, x)$ 。已证明, 核函数  $k(x_i, x)$  只要满足 Mercer 条件即可满足上述要求。

常用的核函数有:

多项式核函数:  $k(x_i, x) = (x_i \times x + 1)^d$

径向基核函数:  $k(x_i, x) = \exp(-g \| x_i - x \|^d)$

感知机核函数:  $k(x_i, x) = \tanh[b(x \times x_i) + c]$

### 3.2 低层特征与高层语义之间的关联

文章使用一种方法, 利用 SVM 的分类特性, 建立低层特征与高层语义的关联来提取语义信息, 并应用于图像的语义检索中。HSV 颜色直方图 72 维、灰度共生分布 4 维, 将它们作为支持向量机的输入向量, 对图像类进行学习, 建立图像低层特征与高层语义的关联。

## 4 实验过程及结果分析

实验环境:

实验的环境是在 PC 机上, 采用 Visual C++ 语言, SQL Server 2000 为数据库, 选取了四类图像: 汽车、花朵、日出、地图。共 1016 张图片, 将图像的相对物理地址存放在数据库中, 方便对图像进行读取。对每类图像进行特征提取, 将提取出来的颜色特征和纹理特征存放在数据库中。

分类器的构造:

选取其中一类图像作为正例, 其它三类图像作为反例, 构造训练集。共 500 个特征向量。100 个正例、300 个反例, 其它图像每种图像含 100 个反例。测试集包括 27 个正例、111 个反例, 其它图像每种图像包括 37 个反例。

在进行训练时, 分别尝试 3.1 节中提到的三中核函数, 并最终选择具有最好实验效果的径向基核函数。这种算法需要建立四个语义分类超平面, 每一个超平面与其它类分隔开。实验过程中进行了三个实验, 实验一单独从数据库中读取 HSV 颜色特征, 共 72 维向

量,使用 SVM 分类器进行分类。实验二单独从数据库中读取纹理特征,共 4 维向量,使用 SVM 分类器进行分类。实验三将两种特征结合起来,共 76 维向量,使用 SVM 分类器进行分类。

表 1、表 2、表 3 分别给出以上三个实验的实验结果。

表 1 HSV 颜色直方图与 Texture 纹理特征相结合的结果

类别	训练集		测试集		
	正样本数	负样本数	正样本数	负样本数	正确率 (%)
汽车	120	360	57	171	82.5
花朵	120	360	57	171	83.8
日出	120	360	57	171	95.6
地图	120	360	57	171	93.4

表 2 HSV 颜色直方图结果

类别	训练集		测试集		
	正样本数	负样本数	正样本数	负样本数	正确率 (%)
汽车	120	360	57	171	77.2
花朵	120	360	57	171	80.7
日出	120	360	57	171	85.5
地图	120	360	57	171	90.2

表 3 Texture 纹理特征结果

类别	训练集		测试集		
	正样本数	负样本数	正样本数	负样本数	正确率 (%)
汽车	120	360	57	171	75
花朵	120	360	57	171	75
日出	120	360	57	171	91.7
地图	120	360	57	171	73.7

由以上的三张表可以看出有些图像,例如日出图像用纹理特征,地图用 HSV 颜色特征就可以达到很高的检索正确率,而汽车和花朵图像则需要将颜色和纹理特征结合起来,才能达到很好的检索效果。且总的来说,将多特征融合起来,能够达到更高的检索正确率。

## 5 结束语

语义检索是基于内容的图像检索的发展方向,如何从低层特征中提取高层语义让检索更接近于用户的需要,成为该领域的研究热点之一。文章采用支持向

量机的方法建立图像的低层特征与高层语义的关联,取得较满意的效果。实验结果表明:

(1)不能采用唯一的特征来描述所有的图像的语义,不同的特征或特征组合才能较好地建立关联;

(2)在描述图像时并非采用的特征越多就越能较好地表示一类图像,即语义关联的效果不一定随着低层特征的复杂程度的上升而变得更好。一般而言,背景简单、训练样本正例较多的图像具有较好的语义关联。

为了能取得更好的效果和更广泛的使用范围,今后的工作中,还需要继续研究和改进,使检索的正确率达到更高。其中一种解决的方法是引入更多的图像特征信息(如形状)及对图像进行分割等;另外,为进一步提高检索的准确率还可以引入相关反馈机制。

## 参考文献:

- [1] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003: 3-5.
- [2] Xie Hua, Omega A. Feature Representation and Compression for Content-based Retrieval[C]//Proceedings of SPIE - The International Society for Optical Engineering. [s. l.]: [s. n.], 2001: 111-122.
- [3] 荆延国. 一个基于图像中语义对象的图文双向查阅系统的设计与实现[D]. 大连: 大连海事大学, 2000: 4-6.
- [4] 王上飞, 陈恩红, 汪祖媛, 等. 基于支持向量机的图像情感语义注释和检索算法的研究[J]. 模式识别与人工智能, 2004, 17(1): 27-33.
- [5] Heckbert P S. Color image quantization for frame buffer display[J]. Comput. Graph, 1982, 16: 297-307.
- [6] Haralick R, Shanmugan K, Dinstein I. Textural features for image classification[J]. IEEE Trans System Man Cybernetics, 1973(3): 610-621.
- [7] 徐建华. 图形处理与分析[M]. 北京: 科学出版社, 1992.
- [8] Manjunath B S, Ma W Y. Texture Features for browsing and retrieval of large image data[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1996, 18(8): 837-842.
- [9] Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines [M]. Cambridge: Cambridge University Press, 2000.
- [10] Gunns. SVM for Classification and Regression[R]. [s. l.]: University of Southampton, Faculty of Engineering Science and Mathematics, School of Electronics and Computer Science, 1998.

中国计算机学会会刊、中国科技核心期刊  
《计算机技术与发展》欢迎订阅, 邮发代号: 52-127