

增量学习语义属性的图像内容检索系统增强*

丁学东, 刘 渊, 谢振平

(江南大学 数字媒体学院, 江苏 无锡 214122)

摘 要: 大规模图像内容检索是实现图像语义信息获取的重要手段,其首要需解决图像低层特征与用户高层语义间的语义鸿沟问题。针对该问题,引入图像语义属性,并结合增量分类学习方法(online core vector machine, OCVM),提出了一种增量构建大规模图像内容检索系统的新方法。该方法借助检索反馈学习机制可以提升图像语义属性的辨别准确性,能在扩张图像库规模的同时,提升图像内容检索的可靠性。实验结果表明了上述方法的有效性,其检索性能可逐步地达到离线构建方法的最佳性能,但具有更好的可扩展性和自提升能力。

关键词: 图像内容检索; 语义属性; 低层特征; 增量学习

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2014)01-0273-04

doi:10.3969/j.issn.1001-3695.2014.01.064

Improvement of content-based image retrieval system by incrementally learning semantic attributes of images

DING Xue-dong, LIU Yuan, XIE Zhen-ping

(School of Digital Medium, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: The large-scale content-based image retrieval system must be an important means of getting semantic information, firstly it needed to solve the problem of the semantic gap which between the high-level semantics the users needed and the low-level features of the image. Facing to this question, it introduced the image semantic attributes, and combined with incremental learning method OCVM, this paper proposed a new method which based on building a new large-scale content-base image retrieval system incrementally. It used the retrieval feedback learning mechanisms to enhance the accuracy on identifying the image semantic attributes. With the expansion of the scale of the image library, it could improve the reliability of the content-based image retrieval at the same time. The experimental results show the effectiveness of the above method, and the retrieval performance of the method can gradually achieve the method which constructed offline, but has better scalability and improving capacity itself.

Key words: content-based image retrieval(CBIR); semantic attributes; low-level features; incrementally learning

传统的图像内容检索(CBIR)通常采用以下三个步骤^[1,2]:

a)根据图像的颜色、纹理、形状等低层特征提取图像的特征矢量,即将图像进行数字化表征;b)利用提取的特征矢量计算图像库中的图像与用户输入图像的相似度;c)通过判断相似度的大小进行图像检索。其问题是,图像的数字化表征难以很好地反映图像语义内容,图像的低层视觉特征与用户所需的高层语义之间存在着语义鸿沟^[3,4]。因此,基于语义的图像检索已成为当前图像检索研究的热点^[5,6]。文献[7]根据图像的颜色、边缘等低层特征与高层语义之间的映射关系,提出了一种层次化图像语义检索思想,并通过 Bayes 分类的方法加以实现。文献[8]采用神经网络的方法,将图像按语义组成目录,采用非层次化的分类进行图像检索。文献[9]提出了使用属性表示图像的方法,并基于图像语义属性进行图像内容检索。

上述几种方法均只考虑了在图像库内容已确定情况下的图像检索系统构建,当需要对原图像数据库进行扩容时,需要将原始图像库和新增图像合并后,完全重新训练构建检索系统。显然,这些研究成果不太适用于实际的图像语义内容检索系统构建。现实中,经常要求能在当前系统基础上实现快速、

高效的扩容,满足系统的可伸缩性。

为此,本文提出一种通过增量学习图像语义属性的图像内容检索系统构建新方法。新方法中,通过引入图像语义的属性分类策略,解决图像低层特征与高层语义间的语义鸿沟问题;同时引入在线核向量机(OCVM)^[10]方法对新增图像进行学习,实现图像检索库的增量扩容。此外,利用 OCVM 的增量反馈学习能力,在扩容图像库的同时,借助用户反馈机制逐步提升检索系统的可靠性。

1 相关背景知识

1.1 图像低层特征

1.1.1 基于 HSV 颜色空间的色彩特征提取

HSV 颜色空间是一种面向视觉感知的颜色空间模型,广泛应用于计算机视觉领域。它把彩色信号表示为色调(hue)、饱和度(saturation)和亮度(value)三种属性。HSV 颜色空间可由 RGB 空间转换而来,能够很好地与人眼的视觉特征相匹配,因此应用这种彩色模型会更适合用户的视觉判断^[11]。本文采用式(1)(2)的量化方法对 H、S、V 三个分量进行非等间隔量化。

收稿日期: 2013-03-18; **修回日期:** 2013-04-23 **基金项目:** 国家自然科学基金资助项目(60975027);无锡市科技支撑计划(社会发展)资助项目(CSE01N1206)

作者简介: 丁学东(1987-),男,内蒙古乌兰察布人,硕士研究生,主要研究方向为图像处理(531404355@qq.com);刘渊(1967-),男,教授,硕士,主要研究方向为计算机网络安全、数字媒体;谢振平(1979-),男,博士,主要研究方向为模式识别、图像处理。

$$H = \begin{cases} 0 & h \in [0, 20) \cup [315, 360] \\ 1 & h \in [20, 40) \\ 2 & h \in [40, 75) \\ 3 & h \in [75, 155) \\ 4 & h \in [155, 190) \\ 5 & h \in [190, 1270) \\ 6 & h \in [270, 295) \\ 7 & h \in [295, 315) \end{cases} \quad (1)$$

$$S = \begin{cases} 0 & s \in [0, 0.2) \\ 1 & s \in [0.2, 0.8) \\ 2 & s \in [0.8, 1] \end{cases}, V = \begin{cases} 0 & v \in [0, 0.2) \\ 1 & v \in [0.2, 0.8) \\ 2 & v \in [0.8, 1] \end{cases} \quad (2)$$

对 H 、 S 、 V 三维特征矢量取不同权重合成一维特征向量:

$$L = HQ_S Q_V + SQ_S + V \quad (3)$$

式(3)中, Q_S 、 Q_V 分别是分量 S 、 V 的量化级数, 这里 $Q_S = Q_V = 3$ 。

因此, 式(3)可变为 $L = 9H + 3S + V$, L 的取值范围为 $[0, 1, \dots, 71]$, 计算 L 获得 72 维的颜色特征向量。

1.1.2 基于 HOG 的形状特征提取

图像中物体的形状可以用其梯度或者边缘方向分布很好地描述。根据这种理论, Dalal 等人^[12]提出了 HOG (histogram of oriented gradients) 算法, 基于 HOG 的形状特征对图像的几何形变能保持很好的不变形。可以采用以下三个步骤得到图像的 HOG 特征: a) 对图像进行规范化, 将输入图像转换为灰度图像, 然后采用 Gamma 校正对输入图像的颜色空间进行标准化, 并利用高斯滤波平滑图像; b) 计算梯度, 使用最简单的一维离散梯度模板 $[-1, 0, 1]$ 对图像上的点 (x, y) 处水平和垂直两个方向进行处理, 进而得到图像在该点的梯度; c) 提取图像的 HOG 特征。

Dalal 等人^[12]将分辨率为 64×128 像素大小的图像进行区域划分: 把 4×4 大小的像素区域设为一个单元格 (cell), 在每个 cell 内进行梯度统计, 梯度方向的范围是 $0 \sim 360^\circ$, 梯度方向直方图分为九级, 计算每个 cell 内对应的梯度幅值加权的梯度方向直方图, 得到一个九维的特征向量; 将 2×2 个 cell 组成一个 8×8 像素的块 (block), 相邻的两个 block 有重叠区域, 大小为原区域的 50%。如图 1 所示, A1、A2、A3、A4、B1、B2、B3、B4 分别为一个 cell, 其中 A1、A2、A3、A4 组成一个 block A; B1、B2、B3、B4 组成另一个 block B; A2、B1、A4、B3 组成区域为 block A 和 B 的重叠区域, 即用橙色表示的区域 (见电子版)。

HOG 特征向量维数的计算方法是: 每个 cell 的维数是 9, 每个 block 有 4 个 cell, 所以一个 block 的维数是 $4 \times 9 = 36$, 检测窗口内共有 $7 \times 15 = 105$ 个 block, 所以一幅图像的 HOG 特征向量维数是 $105 \times 36 = 3780$ 。得到 HOG 特征向量后, 采用式(4)对 HOG 特征向量进行归一化处理。其中 f_i 为原特征向量, f_i^* 为归一化后的特征向量, ε 为常数。通过归一化处理可以进一步去除光照等因素的影响。

$$f_i^* = \frac{f_i}{\sqrt{\sum_{i=1}^N f_i^2 + \varepsilon}} \quad (4)$$

1.2 OCVM 在线分类器

1.2.1 MEB 问题

Tsang 等人^[13]提出核向量机 (core vector machine, CVM) 算法, 该算法将 SVM 求解问题转换为最小球问题 (minimum enclosing ball, MEB), 进而使用 $(1 + \varepsilon)$ 次迭代求解 MEB 问题^[14]。

算法具体为: 给定一个数据集 $S = \{x_1, \dots, x_m\}$, 最小球问题的

目标是找到一个最小超球体 $B(c, R)$, 使其包含所有样本 S , 其中 c 和 R 分别为最小超球的球心和半径。MEB 问题可表示为

$$\min_{R, c} R^2 \quad \text{s. t.} \quad \|c - \varphi(x_i)\|^2 \leq R^2 \quad i = 1, \dots, m \quad (5)$$

利用拉格朗日技巧, 其对偶形式为

$$\max_{\alpha} \alpha^T \text{diag}(K) - \alpha^T K \alpha \quad \text{s. t.} \quad \alpha \geq 0, \alpha^T I = 1 \quad (6)$$

其中: $\alpha = [\alpha_1, \dots, \alpha_m]^T \geq 0$ 为拉格朗日乘子向量, $I = [1, \dots, 1]^T$, $K_{m \times m} = [k(x_i, x_j)]$ 是核矩阵。求解上述 QP 问题, 得到 α 后, 有

$$c = \sum_{i=1}^m \alpha_i \varphi(x_i) \quad (7)$$

$$R = \sqrt{\alpha^T \text{diag}(K) - \alpha^T K \alpha} \quad (8)$$

当 $k(x_i, x_j)$ 为常数时, 原问题可改写为

$$\max_{\alpha} -\alpha^T K \alpha \quad \text{s. t.} \quad \alpha \geq 0, \alpha^T I = 1 \quad (9)$$

1.2.2 CVM 和 OCVM

单分类支持向量机的求解问题可以转换成一类特殊的 MEB 问题, 相应的方法称为 CVM。CVM 算法主要分为以下三个步骤: a) 令 $S_0 = \{\varphi(z_0), c_0 = \varphi(z_0), R_0 = 0, t = 0\}$, 寻找距离 c_i 最远的 $\varphi(z)$ 点, 若该点落在 $B(c_i, (1 + \varepsilon)r)$ 之内, 则 c_i 终止迭代, 否则 $S_{t+1} = S_t \cup \{\varphi(z_t)\}$; b) 求解 MEB(S_{t+1}); c) 令 $t = t + 1$, 返回步骤 a)。

OCVM 算法将 CVM 应用于在线学习中, 通过对每一个新增加的数据进行训练, 若其落于 $B(c_i, (1 + \varepsilon)r)$ 之内, 则取下一个数据继续学习, 否则令 $S_{t+1} = S_t \cup \{\varphi(z_t)\}$, 求解 MEB(S_{t+1})。

2 CBIR 的语义增量构建

2.1 图像语义属性

为了解决图像的低层特征与用户高层语义间的语义鸿沟问题, 本文引入图像语义属性^[9], 用图像语义属性表达图像中所包含的语义信息。将图像库中的图像利用与其相关的语义属性表示, 采用此种方法不仅可以描述熟悉的事物, 而且可以对某些未知的事物进行描述。如图 2 所示, 分别用四种语义属性来描述 bird (鸟) 和 plane (飞机), 第三幅图像为未知, 同样可以将其描述为一种具有 wheel (车轮)、row_wind (排风) 语义属性的物体, 但不具有 head (头)、legs (腿) 语义属性的物体。

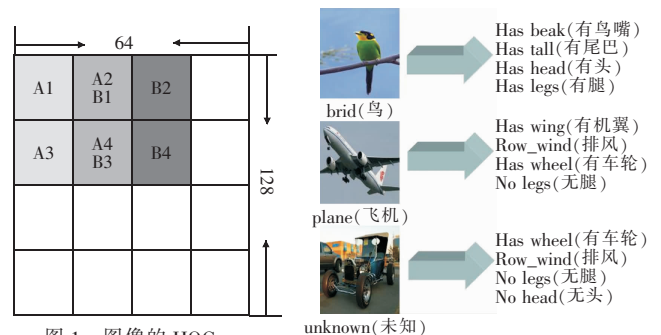


图 1 图像的 HOG 特征区域划分

图 2 图像语义属性描述图像的方法

常采用以下几种图像语义属性命名方法:

a) 根据物体的某个部位命名语义属性, 例如 head (头)、nose (鼻子)、legs (腿) 等。

b)根据物体的形状命名语义属性,例如 2D Boxy(二维正方形)、3D Boxy(三维正方体)等。

c)根据构成物体的材质命名语义属性,例如 wood(木质的)、glass(玻璃)、plastic(塑料的)。

d)根据一些有歧义的语句描述语义属性,例如 horse has but donkey does not have(马具备但是驴不具备的)。

图像语义属性的提取方法如下:a)获取图像语义属性的训练集,将图像库中具有某共同语义属性的图像集作为训练集的正例部分,不具有此语义属性的图像集作为训练集的反例部分;b)利用分类器对训练集进行学习,进而得到图像的语义属性。以获取 head(头)语义属性为例,把 bird(鸟)、horse(马)、donkey(驴)、goat(山羊)等具有 head 语义属性的图像作为训练集的正例,plane(飞机)、bus(公共汽车)等不具有 head 语义属性的图像作为训练集的反例,通过分类器 OCVN 对训练集进行学习,从而得到 head 语义属性。

2.2 系统框架及实现

本文提出的图像检索系统框架结构如图3所示。根据该框架结构得出本文的图像检索方法,主要通过以下五个步骤进行图像检索:

a)从图像库中选择少量图像作为训练集,根据上文介绍的 HSV 的颜色特征和 HOG 形状特征提取图像的低层特征。

b)根据图像语义属性的学习方法,利用 CVM 分类器学习步骤 a)提取的低层特征数据,得到图像的各个语义属性。

c)利用图像的语义属性对待检索的图像库进行分类,根据分类的结果进行图像检索。具体方法为:(a)设图像库中有 N 种图像的语义属性,将图像库中的每一幅图像用这 N 种属性表示;(b)根据图像所对应的属性对图像库进行分类,例如图像库中有 legs(腿)、head(头)、tail(尾巴)、furn_back(毛背)、wing(机翼)、wheel(车轮)六种属性,经过分类器分类后,某一幅图像对应的属性情况为 has legs、has head、has tail、furn_back、no wind、no wheel,则判定该图像为 horse 这一类;(c)用户输入待查询的示例图像,提取示例图像的低层特征,利用分类器判断示例图片类别,根据步骤(b)中图像库的分类结果进行图像检索。

d)用户根据检索结果随机反馈 N 幅图像,提取这 N 幅图像的低层特征,利用在线分类器 OCVN 学习图像低层特征,从而实现对图像语义属性的增量学习。在该步骤中,因为 OCVN 分类器可以记录 CVM 分类器对图像数据库已完成的分类信息,所以仅需要训练用户反馈的 N 幅图像就可以得到新的图像库语义属性,该语义属性在辨别图像准确率方面要优于原图像语义属性。

e)重复步骤 c)d)直到用户不再反馈为止。

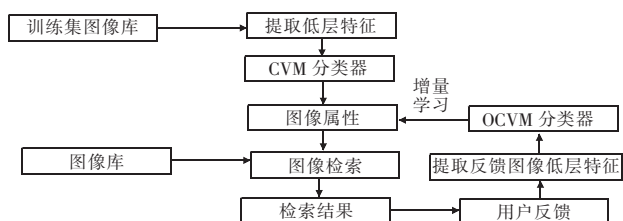


图3 图像检索系统框架结构

从以上方法可以看出,本文的图像检索系统通过引入图像语义属性,结合增量学习的方法,并借助检索反馈学习机制有效地提升了图像语义属性的辨别准确性,能在扩张图像库规模的同时,提升图像内容检索的可靠性,具有较好的实用性。

3 实验及结果分析

3.1 实验设置

本文实验选用两个不同数量规模的实验图像库,其中一个图像库含 5 000 幅图像,定义为图像库 1,另一个图像库是 Corel 图像库,图像数量为 1 000,定义为图像库 2。图像库 1 由 Yahoo 图像库(共 2 237 幅图像)、部分的 PASCAL 图像库以及百度图像库三部分组成。本文从 PASCAL 数据集^[15]选取了 10 种语义属性用于图像检索:legs(腿)、head(头)、beak(鸟嘴)、tail(尾巴)、furn_back(毛背)、wing(机翼)、side_mirror(侧镜)、wheel(车轮)、row wind(排风)、flower(花)。根据这 10 种属性可以检索得到五个种类:bus(公共汽车)、plane(飞机)、horse(马)、bird(鸟)、flower(花)。表1显示了图像库中的两种图像低层特征、十种语义属性和五个种类的对对应关系。

本文语义属性训练集的图像分别选自图像库 1 和图像库 2,flower 语义属性训练集的图像选自图像库 2,其他图像语义属性训练集的图像都选自图像库 1。Flower 语义属性训练集的图像数量为 150;side_mirror、wheel、wing、row_wind 语义属性训练集的图像数量是 300;legs、head、tail、furn_back 语义属性训练集的图像数量是 500。

表1 种类、语义属性和低层特征的对对应关系

种类	语义属性	低层特征
bus	side_mirror、wheel、row_wind	HOG
plane	wing、wheel、row_wind	HOG
horse	legs、head、tail、furn_back	HOG
bird	legs、head、beak、tail	HOG
flower	flower	HSV

3.2 实验结果

查准率和查全率是评价图像检索系统性能的重要指标。设 n 表示测试集中总的图像数量,检索结果中共有 m 幅图像, m 幅图像中与要查询的图像属于相同类别的图像数量为 x , n 幅图像中共有 y 幅图像与要查询图像属于相同的类别,则

图像的查准率: $\text{precision} = x/m$

图像的查全率: $\text{recall} = x/y$

本文通过实验验证了 CVM、OCVM 分类器在参数 $C = 1000$, $\varepsilon = 10^{-6}$ 时图像检索系统的查准率最高。其中: C 为 OCVN 的惩罚因子, ε 是用于求解 MEB 问题的参数。分别计算 plane、bus、bird、horse、flower 这五个种类的查准率和查全率,根据每次得到的检索结果随机选取 10~30 幅图像进行反馈,当准确率为 1 或者准确率不再随着反馈次数的增加发生变化时停止反馈。图4为对五个种类(plane、bus、horse、bird、flower)进行反馈后得到的图像检索系统的查准率折线图。图5是与图4相对应的图像检索系统的查全率折线图。

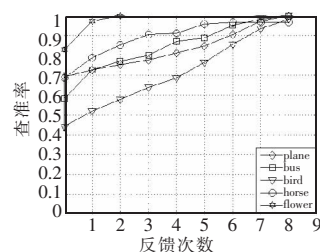


图4 图像检索系统的查准率

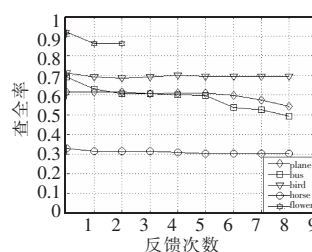


图5 图像检索系统的查全率

图4中,图像检索系统的查准率随着反馈次数的增加呈递增趋势,其中 plane、bus、flower 这三个类的最终查准率为 1, bird 和 horse 的查准率分别保持在 0.981 4 和 0.965 5。从图4

