

基于词汇树层次语义模型的图像检索算法

张月辉, 吴 健, 陆姗姗, 崔志明

(苏州大学 智能信息处理及应用研究所, 江苏 苏州 215006)

摘 要: 解决语义鸿沟必须建立图像低层特征到高层语义的映射, 针对此问题, 本文提出了一种基于词汇树层次语义模型的图像检索方法. 首先提取图像包含颜色信息的 SIFT 特征来构造图像库的特征词汇树, 生成描述图像视觉信息的视觉词汇, 并在此基础上利用 Bayesian 决策理论实现视觉词汇到语义主题信息的映射, 进而构造了一个层次语义模型, 并在此模型基础上完成了基于内容的语义图像检索算法. 通过检索过程中用户的相关反馈, 不仅可以加入正反馈图像扩展图像查询库, 同时能够修正高层语义映射. 实验结果表明, 基于该模型的图像检索算法性能稳定, 并且随着反馈次数的增加, 检索效果明显提升.

关键词: 词汇树; 语义主题信息; 层次语义模型; 语义映射; 图像检索

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)11-0172-05

Algorithms of Vocabulary Tree Hierarchical Semantic Model Image Retrieval

ZHANG Yue-hui, WU Jian, LU Shan-shan, CUI Zhi-ming

(Institute of Intelligent Information Processing and Application, Soochow University, Suzhou 215006, China)

Abstract: Creating a semantic mapping method is important in solving semantic map. This paper proposes a vocabulary tree hierarchical semantic model image retrieval method. First extracting the characteristics of the image, such as the color information, SIFT feature, and then generating the visual vocabulary to describe the image of visual information. Using Bayesian theory map the visual vocabulary to the semantic subject information, then constructing a hierarchical semantic model. So based on this model we realize a semantic content-based image retrieval algorithms. Through the user's relevance feedback, not only can be added to the positive feedback image, but also can be able to amend the high-level semantic mapping. Results perform stability, and with the increase in the number of feedback retrieval results can improve significantly.

Key words: vocabulary tree; semantic subject information; hierarchical semantic model; semantic mapping; image retrieval

1 引言

随着因特网和多媒体技术的迅速发展, 从 20 世纪 90 年代初至今, 基于内容的图像检索 (CBIR) 得到了广泛的研究. 但由于其使用图像纹理、颜色和形状等低层特征作为索引来对图像进行检索^[1], 而图像低层特征表达出的信息与用户对这些特征图像理

解的含义存在不一致性, 即语义鸿沟, 所以检索效果往往不能满足用户的需求.

如果检索过程中能够考虑图像所包含的图像高层语义信息, 加入用户对图像内容的理解, 则会是一种更贴近用户理解的图像检索. 这种在获取图像语义概念基础上的语义图像检索受到了学者的广泛关注. 语义图像检索过程可归结为 3 个方面: (1) 图像

收稿日期: 2012-05-20; **修回日期:** 2012-06-22

基金项目: 国家自然科学基金 (60970015, 61003054, 61170020); 江苏省高校自然科学基金项目 (10KJB520018); 苏州市科技支撑计划项目 (SG201257)

低层特征提取;(2)图像高层语义映射;(3)语义检索算法的实现.在语义检索过程中,图像高层语义映射的实现是最重要的一步,研究者通过在各方面研究,总结出高层语义映射方式主要有以下几种方式^[2]:(1)人工标注图像的文本信息;(2)利用图像分割和提取的低层特征,得到图像中不同对象的特征描述,并且利用对象之间的空间关系或者特定场景下的某些特性来得到图像的语义信息;(3)提取图像的全局特征,对图像进行聚类学习,并在此基础上关联图像和语义进行建模^[3].

在国内外已有一些专家学者在语义图像检索方面做出了重要成果.如文献[4]中 Vailaya 在利用 2 类贝叶斯分类器将度假图像进行简单的分类:首先判断一个图像是室内或者室外,然后分别对各类图像再进行简单的分类,这些 2 类 Bayesian 分类器组合起来,对于简单的场景分类,有很好的效果.在文献[5]中 Blei 等使用稍微复杂的 CORR-LDA 模型为关键词和图像创建一个基于语言的关联,从而产生图像的语义标注.在文献[6]中 Feng 等提出多贝努里相关模型,利用训练集计算关键词和图像特征的联合概率分布,使用 MBRM 估计关键词概率,使用无参数核密度函数估计图像区域特征的概率,然后使用训练好的模型对测试集进行标注.

图像高层语义特征需要人为理解,而每个用户对某一幅图像的语义信息理解又不尽相同,所以适当的用户反馈信息能够很好的提高检索效果.目前语义检索中语义映射关系大多是通过人工标注建立^[7],但是建立的映射关系不能随着用户检索的反馈信息自动进行修正,从而影响检索效果的提升.针对此问题,本文提出了一种基于词汇树层次语义模型的图像检索方法.首先提取图像包含颜色信息的 SIFT^[8-9]特征来构造图像库的特征词汇树,生成描述图像视觉信息的视觉词汇.并在此基础上利用 Bayesian 决策理论实现视觉词汇到语义主题信息的映射,进而构造了一个层次语义模型,并在此模型基础上完成了基于内容的语义图像检索算法.所以本文在检索过程加入了用户相关反馈.通过检索过程中用户的相关反馈,不仅可以加入正反馈图像扩展图像查询库,同时能够修正高层语义映射,使检索效果更符合用户的查询意图.

2 基于 SIFT 词汇树的层次语义模型

本章将首先介绍基于 SIFT 词汇树的层次语义模型整体框架,然后分别介绍该模型中视觉词汇层

和高层主题语义信息层的构造过程,最后给出视觉词汇层到高层主题语义信息层的映射方法.

基于 SIFT 词汇树的层次语义模型如图 1 所示.

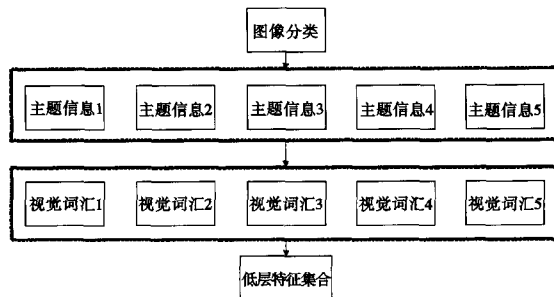


图1 基于 SIFT 词汇树的层次语义模型

首先我们将图像库所有图像提取包含颜色信息的 SIFT 特征作为模型的输入,然后对这些特征向量构造视觉词汇树,将每幅图像量化到生成的视觉词汇节点,这样就构成了层次语义模型的视觉词汇层,在此基础上利用 Bayesian 统计学习模型实现视觉词汇层到高层主题语义信息层的映射.此处主题语义层是根据图像库中的人工标注信息抽取出的语义主题,如野人、汽车等.

2.1 SIFT 词汇树构造

在图像检索中,比较常用的低层特征包括图像颜色、纹理和形状特征. SIFT 特征是一种图像局部纹理特征,提取图像 SIFT 特征时首先利用 DOG 图像空间提取图像中具有尺度不变的关键点,然后对关键点进行相应处理使其具有旋转不变性和一定的光照不变性,生成最终的特征描述符.本文中我们在 SIFT 特征中加入了颜色信息,更好的描述了图像特征.

构造词汇树时,首先提取图像库中所有图片的 SIFT 特征,得到一个特征集合 $\text{Feature} = \{\text{feature}_i\}$,然后对特征集合 $\text{Feature} = \{\text{feature}_i\}$ 进行分层聚类,此处我们采用 K-Means 聚类方法,因为它有比较好的伸缩性,聚类速度相对较快.聚类结果生成 K 个子节点(本文中 K 取 6),计算出每个子节点的中心向量 C_i ,同样地,对新生成的每个子节点再进行 K-Means 聚类,以生成下一层的子节点,不断重复直到树的深度达到我们的预设值 P (本文中 $P=4$).这样最终会构造出一个完整的 SIFT 词汇树.该过程如图 2 所示.

这样通过构造词汇树,就生成了 M 个聚类簇

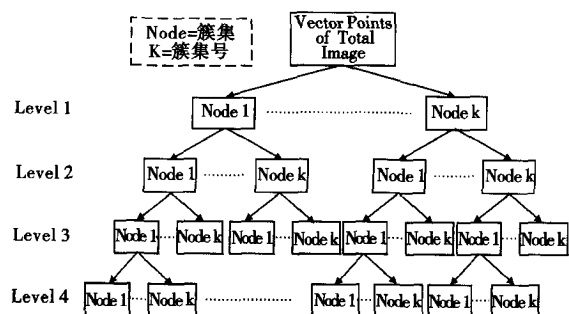


图2 构造词汇树过程

集,且每个聚类簇集都有一个中心向量 C_m 。仿照文本检索中的概念,我们把生成的聚类簇集称作视觉词汇单词,图像库中的每幅图片称为一个文档,通过计算每个文档和词汇单词的相关程度使图像量化到这 M 个聚类簇集中,即生成一个 M 维的向量。假设 F_i 表示每个图片文档, W_m 表示每个词汇单词, $\text{Weight}(i, m)$ 表示文档和词汇单词的相关度。所有每个文档就可以量化到 M 个视觉词汇单词中,在计算 $\text{Weight}(i, m)$ 时,我们使用 tf-idf 语词加权方案。假设 NUM_{im} 表示单词 W_m 在图片文档 F_i 中出现的次数。 N_m 表示词汇单词 W_m 一共包含的图片文档数目,定义:

$$\text{idf}_m = \lg \frac{N}{N_m} \quad (1)$$

最终图片文档 F_i 和视觉词汇单词 W_m 的相关度定义为

$$\text{Weight}(i, m) = \text{NUM}_{im} \times \text{idf}_m \quad (2)$$

通过构造 SIFT 词汇树,生成的 M 个视觉词汇节点就构成了层次语义模型中的视觉词汇层。在检索过程中,每个查询图片也利用相同方法量化到视觉词汇单词上。

2.2 视觉词汇层到语义信息层的映射

本节主要介绍如何利用 Bayesian 统计决策理论实现从视觉词汇层到语义信息层的映射。如上文所述,我们根据图像库中的人工标注信息抽取出语义主题,如野人、汽车等,语义信息层就是由这些主题信息构成,从而每幅图像都能够用这 q 个主题信息来描述。

Bayesian 统计学习模型是一种基于概率的不确定性推理方法。已成功应用于专家系统等领域中。其为图像分类识别问题提供了一种有效的学习框架。Bayesian 理论主要是在获得先验类属概率的前提下,通过对后验概率函数的计算进行分类,计算公式如下:

$$P(q_i, y) = \frac{f_Y(q_i, y) \cdot P(q_i)}{f_Y(y)} \quad (3)$$

将其应用到本文中,可以实现从视觉词汇层到语义信息层的映射。其中, $P(q_i, y)$ 为视觉词汇单词 M_y 映射到主题信息 q_i 的后验概率, $P(q_i)$ 为主题语义分类信息的先验概率值,计算公式如下:

$$P(q_i) = \frac{N_{q_i}}{N} \quad (4)$$

其中 N_{q_i} 是主题信息节点 q_i 包含的图片数, N 是图像库中的总图片数。式(3)中 $f_Y(q_i, y)$ 是视觉词汇单词节点的中心向量 y 相对于主题信息节点 q_i 的条件概率密度,此处我们采用 Parzen 窗方法^[10],对主题信息节点 q_i 对应的每一个词汇单词(即包含 q_i 个节点的主题信息的图像所分布的每一个视觉词汇单词节点),进行同一 Gaussian 核函数求和,近似得到视觉词汇单词节点的中心向量 y 相对于主题信息节点 q_i 的条件概率密度函数如下:

$$f_Y(q_i, y) = \frac{1}{N_{q_i}} \sum_{j \in q_i} G(y, y_j, \delta) \quad (5)$$

其中, $G(y, y_j, \delta)$ 是 Gaussian 核函数, δ 是平滑参数。将公式(4)(5)代入公式(3)中,得到后验概率:

$$P(q_i, y) = \frac{\sum_i G(y, y_j, \delta) \cdot f(y_i)}{\sum_i G(y, y_j, \delta)} \quad (6)$$

其中, $f(y_i)$ 是中心向量 y 的视觉词汇单词上出现包含语义主题信息节点 q_i 的图片数目与总的图像库中数目的比值。这样利用 Bayesian 统计决策理论实现了视觉词汇层到语义信息层的映射,从而实现了图像的高层语义信息提取,并且在此基础上可以进行语义图像检索。

3 检索算法和相关反馈

利用建立好的基于 SIFT 词汇树的层次语义模型,使用多层检索算法,可以完成基于内容的图像检索系统。检索过程中利用视觉词汇层和主题信息层结合进行检索,其算法过程描述如图3所示。

3.1 检索过程

(1) 对待查询图片提取包含颜色信息的 SIFT 特征,得到具有 n 个特征点的特征向量 $Y = \{y_1, y_2, \dots, y_n\}$ 。

(2) 将查询图片的 SIFT 特征量化到 SIFT 词汇树的 m 个视觉词汇单词上,形成一个 m 维的特征向量 $V = \{v_1, v_2, \dots, v_m\}$ 。具体方法是对查询图片提取到的 SIFT 特征点向量对词汇树逐层的节点进行欧

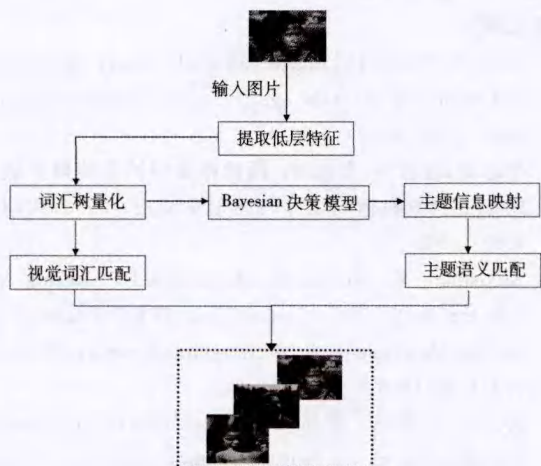


图3 检索算法描述图

氏距离计算,找到每层上最相近的节点作为匹配节点.对提取到得 n 个特征点进行相同操作,最终对于每一个视觉词汇单词节点,累计查询图片中特征点与该节点的匹配次数 NUM_m . 利用生成词汇树时产生的视觉词汇单词节点的 idf_m 值,就可以计算出查询图片与视觉词汇单词节点的相关程度 w_m . 最终形成一个 m 维的特征向量 $L = \{l_1, l_2 \cdots l_m\}$.

(3) 利用在视觉词汇层上量化后的特征向量 $L = \{l_1, l_2 \cdots l_m\}$,对图像库中所有图片在视觉词汇层上量化结果进行匹配,这样就会和每张图片得到一个匹配值 Wt_1 .

(4) 根据(2)步骤中计算得到的 NUM_m ,利用所建立的主题信息语义映射的 Bayesian 模型,按照式(6)计算得到主题语义信息集合对应的概率值集合 $q = \{q_1, q_2 \cdots q_n\}$.

(5) 对于步骤(4)中生成的集合 $q = \{q_1, q_2 \cdots q_n\}$,图像库中图片如果包含主题信息 Q_i ,则该图片和查询图片的语义相关度就等于主题信息 Q_i 在 $q = \{q_1, q_2 \cdots q_n\}$ 中的概率值,记为 Wt_2 .

(6) 最终图像库中的图片 IMG_j 和查询图片的相似度定义 $Weight = Wt_1 \times Wt_2$.

3.2 相关反馈

在上述基于词汇树的语义模型下进行图像检索时,视觉词汇层到主题语义信息层的映射对检索效果有很重要的影响.所以在检索过程中,可以根据用户搜索结果的反馈信息修正语义映射^[11]使其更加完善,并且可以加入正反馈图像来扩展图像库,因此避免了人工标注图像语义信息来增加查询图像库的困难.具体过程如下.

(1) 根据用户的查询图片返回查询结果,用户

根据返回结果标记满意的结果为正反馈记录 I^+ ,其余为负反馈记录 I^- .

(2) 将用户的查询图片加入到词汇树中,其量化到的每一个视觉词汇单词节点,都加入该图片的记录,并且更新 idf 值.

(3) 根据语义映射公式(6)在视觉词汇层到主题语义信息层映射的过程中,对于正反馈图像分布的每一个视觉词汇单词节点,其视觉词汇单词 y 上出现包含语义主题信息节点 q_i 的图片数目占总图像库中数目的比值 $f(y_i)$ 都增加一个正比例因子 U^+ ,而对于负反馈图像分布的每一个视觉词汇单词节点,其 $f(y_i)$ 都加上一个负比例因子 U^- .

这样在反馈过程中,通过用户提供正反馈图像和负反馈图像,不仅可以自动扩展含有高层语义信息的查询图像库,而且使得建立的基于词汇树层次语义模型通过修正视觉词汇到主题信息的映射关系,使图像的语义信息提取更加准确,随着用户查询次数的增加,视觉词汇到主题信息的映射会越来越完善,检索效果更加精确.

4 实验结果及结论

为了验证上述基于词汇树层次语义模型的图像检索效果,我们选取标准 corel 库中 1000 张图片作为图像检索库,根据其语义标注信息抽取出主题语义信息为野人、海滩、建筑、汽车、恐龙、大象、鲜花、骏马、雪山、美食共 10 类,每一类主题语义信息包含图片 100 张.提取这 1000 张图片的包含颜色信息的 SIFT 特征,构造 SIFT 词汇树,并在此基础上完成主题语义信息映射.并在此语义模型上进行图像检索.

从图 4 和图 5 的实验结果中可以看出,当用户



图4 交互之前的查询结果



图5 交互之后的查询结果

查询一张公交车图像时,在没有用户反馈的时候,查询结果中返回了 32 张相似的公交车图像,检索效果不尽如人意,但是当用户经过了若干次相关反馈之后,查询结果中返回了 62 张相似的公交车图像,较开始的查询结果有很好的提升.图 6 是在每个语义类别中,随着用户反馈次数的增加,检索效果的变化情况.其表明在基于词汇树层次语义模型的图像检索系统中,用户的相关反馈对于检索结果提升有很好的作用.并且随着用户反馈次数的增加,检索出来的相关图像会越多,效果越好.

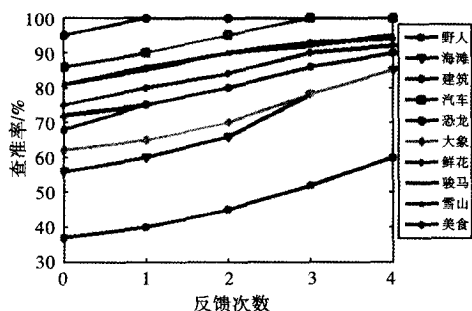


图6 相关反馈

综上所述,本文提出的基于词汇树层次语义模型的图像检索方法是行之有效的,查询结果能够很好的反应用户对图像的主观理解,而且随着用户反馈次数的增加,检索效果有很好的提升.

参考文献:

- [1] Datta R, Joshi D. Image retrieval: ideas, influences, and trends of the new age[J]. ACM Computing Surveys. 2008,40(2):5-60.
- [2] 李志欣,施智平,李志清. 图像检索中语义映射方法综述[J]. 计算机辅助设计与图形学学报, 2008,20(8): 1085-1096.
- [3] Moosmann F, Nowak E. Randomized clustering forests for image classification[J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 2008, 30(9):1632-1646.
- [4] Aditya Vailaya. A bayesian framework for semantic classification of outdoor vacation images[C]// Proceedings of SPIE: Storage and Retrieval for Image and Video Databases. San Jose CA:IEEE, 1999:415-426.
- [5] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research. 2003,3(1):993-1022.
- [6] Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington D C:IEEE, 2004:1002-1009.
- [7] Douillard B, Fox D. Classification and semantic mapping of urban environments[J]. International Journal of Robotics Research. 2011,30(1):5-32.
- [8] Luo Juan. A comparison of sift, pca-sift and surf [J]. International Journal of Image Processing (IJIP). 2009,3(4):187-245.
- [9] 陈赞,沈一帆. 基于词汇树的图片搜索[J]. 计算机工程, 2010,36(6):189-195.
- [10] Zhou Xiang-jun, Zhou Ding-xuan. High order Parzen windows and randomized sampling[J]. Advances in Computational Mathematics. 2009,34(4):68-349.
- [11] 苏小红,丁进,马培军. 用兴趣点凸包和 SVM 加权反馈实现图像检索[J]. 计算机学报, 2009,32(11): 2221-2228.

作者简介:

张月辉 男,(1987),硕士研究生.研究方向为智能信息处理.