

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2013 **3**
Vol.18 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2013 年 3 月 第 18 卷 第 3 期(总第 203 期)

目 次

图像处理和编码

沿边局部灰度差分椒盐噪声滤波 孔勇奇, 潘志庚(249)

图像分析和识别

核稀疏保留投影及在步态识别中的应用 王科俊, 阎涛, 吕卓纹, 唐墨(257)

局部 Gist 特征匹配核的场景分类 杨昭, 高隼, 谢昭, 吴克伟(264)

图像理解和计算机视觉

条件随机场模型的场景描述 赵龙, 郭立, 谢锦生(271)

基于 L_1 与空间点分类的 3 维场景重建 李静, 杨宜民, 付根平, 陈建平(277)

稀疏表示的 Lucas-Kanade 目标跟踪 徐如意, 陈靛影(283)

医学图像处理

管状特性和主动轮廓的 3 维血管自动提取 尚岩峰, 汪辉, 汪宁, 陈杰, 姚丽萍, 孙琨(290)

“第 17 届 CAD&CG 学术会议”专栏

多步攻击的规则树检测及可视化 胡亮, 赵剑明, 解男男, 努尔布力(299)

显著性边缘引导下的图像抽象化 王勋, 李红, 刘春晓, 严彩萍(305)

稀疏表达的运动数据压缩 齐天, 肖俊, 庄越挺(311)

基于语义的用户关注度计算 陈蕾英, 陈佳舟, 潘斌, 彭群生(318)

九宫格空间框架的图书图像检索 李宗民, 唐志辉(325)

光学相干层析视网膜体数据的 3 维分割 樊鲁杰, 孙延奎, 张田, 田小林(330)

混合软影绘制算法 刘浏, 周炜, 李华(336)

基于场的人群运动仿真 赵欣欣, 张勇, 孔德慧, 尹宝才(344)

利用大位移视图的自动可信图像修补 刘春晓, 金剑秋, 彭群生(351)

第八届国际数字地球会议征稿通知 (358)

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 18 No. 3 March 2013

Contents

Image Processing and Coding

Salt and pepper noise filtering algorithm based on local border gray-scale differences Kong Yongqi, Pan Zhigeng(249)

Image Analysis and Recognition

Kernel sparsity preserving projections and its application to gait recognition Wang Kejun, Yan Tao, Lv Zhuowen, Tang Mo(257)

Scene categorization of local Gist feature match kernel Yang Zhao, Gao Jun, Xie Zhao, Wu Kewei(264)

Image Understanding and Computer Vision

Scene description based on the conditional random fields model Zhao Long, Guo Li, Xie Jinsheng(271)

Three dimensional scene reconstruction based on L_1 and scene point classification
..... Li Jing, Yang Yimin, Fu Genping Chen Jianping(277)

Lucas-Kanade tracking based on sparse representation Xu Ruyi, Chen Jingying(283)

Medical Image Processing

Three dimensional vessel extraction model based on tubular characters and active contour model
..... Shang Yanfeng, Wang Hui, Wang Ning, Chen Jie, Yao Liping, Sun Kun(290)

Issum of the CAD&CG'2012

Multi-step attacks detected by rules tree and visualization Hu Liang, Zhao Jianming, Xie Nannan, Nurbol(299)

Salient edge guided image abstraction optimization Wang Xun, Li Hong, Liu Chunxiao, Yan Caiping(305)

Motion data compression using sparse representation Qi Tian, Xiao Jun, Zhuang Yueting(311)

Semantic-based user interests computation Chen Leiying, Chen Jiazhou, Pan Bin, Peng Qunsheng(318)

Books image retrieval based on grid space frame Li Zongmin, Tang Zhihui(325)

Three dimensional segmentation to detect retinal boundary surfaces from OCT volume data
..... Fan Lujie, Sun Yankui, Zhang Tian, Tian Xiaolin(330)

Hybrid soft shadow rendering method Liu Liu, Zhou Wei, Li Hua(336)

Field-based crowd simulation Zhao Xinxin, Zhang Yong, Kong Dehui, Yin Baocai(344)

Automatic and accurate image completion from a large displacement view Liu Chunxiao, Jin Jianqiu, Peng Qunsheng(351)

中图法分类号: TP39 文献标识码: A 文章编号: 1006-8961(2013)03-0325-05

论文引用格式: 李宗民, 唐志辉. 九宫格空间框架的图书图像检索[J]. 中国图象图形学报, 2013, 18(3): 325-329.

九宫格空间框架的图书图像检索

李宗民, 唐志辉

中国石油大学(华东)计算机与通信工程学院, 青岛 266580

摘要: 图书图像检索是基于内容的图像检索(CBIR)的一个重要应用。尺度不变特征变换(SIFT)与方向梯度直方图(HOG)在CBIR和物体识别等领域有着广泛应用。根据图书图像的特点,把图书图像按照九宫格进行空间划分,并以SIFT特征为基础对每个分块子图像进行特征描述,然后结合HOG特征组成联合特征描述,联合特征对图书图像的描述力较强而且数据量较小。采用“倒数比”的方式进行联合特征相似性度量能够有效防止子图像发生突变而造成的巨大误差。在规模为50 000幅的图书图像库上的实验结果表明,检索的精度和耗时都在可接受范围内。

关键词: 图书图像检索;尺度不变特征变换;方向梯度直方图;九宫格空间框架

Books image retrieval based on grid space frame

Li Zongmin, Tang Zhihui

College of Computer & Communication Engineering, China University of Petroleum, TsingTao 266580, China

Abstract: Books image retrieval is an important application of CBIR (content-based image retrieval). SIFT (scale invariant feature transform) and HOG (histogram of oriented gradient) are widely used in CBIR and object recognition. A book image is divided into nine blocks called sub-images according to its characteristics. We extract features from the image based on SIFT, and then we combine the features with the HOG feature to create a combined feature. The combined feature occupies little data and describes an image well. We use a new way to calculate the combined features' similarity, which will avoid large errors caused by some sub-image. An image set with 50000 book images is used to test our method and we find the accuracy to be acceptable and the time consumed by the retrieval process is quite short.

Key words: books image retrieval; scale-invariant feature transform; histogram of oriented gradient; grid space frame

0 引言

随着互联网与人们生活联系日益密切和电子商务的迅速发展,越来越多的人开始在互联网上搜索商品和商品图片,简单的基于文字标签的搜索已经不能满足人们的需求。基于内容的图像检索^[1]开始受到更多人的关注,并吸引了更多的研究者对

其进行研究,且有些相关应用和产品已经出现在互联网上。手机等移动终端发展迅速,简单方便的拍照搜图操作受到人们的青睐。图书是互联网上被搜索比较多的商品,而且图书量大,种类繁多,对“以图搜图”技术的需求非常迫切。

图像的局部特征描述首先提取图像的局部关键点,然后对关键点进行特征描述。局部关键点描述子主要有SIFT(scale invariant feature transform)^[2-3]、

收稿日期:2012-08-27;修回日期:2012-12-12

基金项目: 山东省自然科学基金项目(ZR2009GL014);山东省中青年科学家奖励基金项目(BS2010DX037);文化部科技创新基金项目(46-2010);中央高校基本科研基金项目(09CX04044A, 10CX04043A, 10CX04014B, 11CX04053A, 11CX06086A, 12CX06083A, 12CX06086A)

第一作者简介: 李宗民(1956—),男,教授,2005年于中国科学院计算技术研究所计算机科学与技术专业获博士学位,主要研究方向为计算机图形学、图像处理、科学计算可视化、计算机辅助几何设计、模式识别等。E-mail: lizongmin@upc.edu.cn

SURF (speeded up robust features)^[4] 以及 PCA-SIFT^[5] 等。其中 SIFT 特征信息量丰富,并且具有旋转、缩放不变形,具有一定的仿射不变性,能适应一定程度的视点变换和光照变化,具有较好的鲁棒性。SURF 特征的鲁棒性不及 SIFT。PCA-SIFT 对 SIFT 特征进行了简化,描述能力有所降低。但是由于 SIFT 特征的数据量过大,即便使用 BOW (bag of words)^[6] 对 SIFT 特征进行量化,要达到精确检索所需数据量仍然非常大,特征匹配难以达到实时性。把图书图像做了空间构架划分,并在此基础上提取了 SIFT 特征的部分有效信息组成新的特征向量,有效地降低了数据量而依然能保持较好的描述能力。HOG (histogram of oriented gradient)^[7] 是一种局部区域描述符,它能够很好地描述物体的形状和边缘,不同的图书图像中文字和图案的形状和边缘信息差别比较大,使用 HOG 特征描述具有很好的效果。

图书大多是矩形,并且图书文字和图案的分布一般比较有规律。如图 1 所示把图书按照九宫格空间框架划分能够使每一个子空间的特征区分度比较高,在每一个子图像上基于 SIFT 特征提取信息形成新的特征,结合 HOG 特征,采用“倒数比”方式组合分块特征进行相似性度量,最终实现图书图像检索。



图 1 图书图像九宫格划分

Fig. 1 Space frame of books images

1 图书图像检索流程

图书检索的目的是在图书库中搜索拍摄图像中的图书,也就是说在图书图像库中寻找与拍摄图像视觉最相似的图像。把图书图像的视觉特征表示成一个特征向量,以两个向量的相似度代表

相应图像的相似度,通过比较特征向量寻找最相似的图书。

图书拍摄图像的检索流程如图 2 所示。当输入拍摄图像时,首先按照九宫格对图像进行划分得到一系列子图像,提取每一个子图像的特征和 HOG 特征,基于 SIFT 特征提取相关信息组成新的描述特征,计算拍摄图像的联合特征并与库图像特征进行比较计算相似度,按照相似度对库图像进行排序,选择相似度最高的若干幅图像作为检索结果输出。

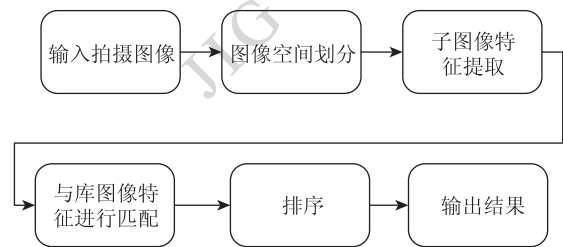


图 2 图书图像检索流程

Fig. 2 The books image retrieval process

基于九宫格的检索流程有以下优点:

1) 图书图像按照九宫格划分后,每一个子图像都具有比较明显的特点,相应的子图像进行匹配后,再把子图像的相似度组合形成新的相似度,使联合特征既有局部性又有全局性。

2) 对每一个子图像的描述可以使用 SIFT 特征数据的统计信息,而避免了直接使用 SIFT 特征,从而大幅度降低了数据量,提高匹配的速度并且降低了内存空间的占用。

3) 可以对不同区域的子图像设置不同的置信度,排除对检索结果有不良影响的子图像,从而提高检索的精度。

2 图书图像特征提取

基于九宫格,在 SIFT 特征基础上提取图书图像特征,并且与 HOG 特征相结合形成了图书图像的联合特征。

2.1 九宫格划分

分析图书图像的特点,发现把图书图像平均划分为 3×3 块形成九宫格,每一个子图像的特征比较明显,如图 3 所示。对于大部分图书,书名大都分布在 4、5、6 区域,少部分分布在 1、2、3 区域,图书的主要图案分布在 7、8、9 区域。

1	2	3
4	5	6
7	8	9

图3 九宫格示意图

Fig. 3 Space frame

2.2 特征选择

SIFT 算法由 Lowe 于 1999 年提出并于 2004 年完善总结^[2-3],是一种局部关键点检测和局部特征提取的有效算法。SIFT 特征信息量丰富,并且具有缩放、旋转不变形,具有一定的仿射不变性,能适应一定程度的视点变换和光照变化。SIFT 算法的主要步骤是:检测尺度空间极值点、定位极值点、确定关键点的主方向,生成特征点描述子。SIFT 特征能够很好地描述图书图像的关键点和信息。

每一个 SIFT 特征点的描述子是 128 维,每一幅图书图像通常能提出上千个特征点,即便使用 BOW 对 SIFT 特征聚类描述图书图像,要想达到精确检索所需的数据量仍然非常大,很难满足特征匹配的实时性,因此对 SIFT 特征进行一定的简化是有必要的。以 SIFT 特征为基础提取了图像的两个特征:

- 1) SIFT 特征点密度。对子图像 $i (i = 1, 2, \dots, 9)$ 计算 SIFT 特征点密度值

$$d_i = \frac{p_i}{s} \quad (1)$$

$$s = \sum_{i=1}^9 p_i \quad (2)$$

式中, p_i 是子图像 i 中 SIFT 特征点数, s 是原图像 SIFT 特征点总数。

这样一幅图像的 SIFT 点密度特征向量表示为 $D = (d_i)$ 。

- 2) SIFT 主方向角度直方图。对子图像 $i (i = 1, 2, \dots, 9)$, 统计 SIFT 特征点主方向的角度直方图 $A_i = (A_{ij})$, $(j = 1, 2, \dots, 18)$, 以 20° 为步长把角度空间分为 18 个 bin, 所以每个子图像得到一个 18 维特征向量。

$$A_{ij} = \frac{p_{ij}}{p_i} \quad (3)$$

$$p_i = \sum_{j=1}^{18} p_{ij} \quad (4)$$

式中, p_{ij} 是子图像 i 中落入第 j 个 bin 的特征点个

数。 p_i 是子图像 i 的特征点总数。

HOG 特征是一种局部区域描述符,它通过计算局部区域上的梯度方向直方图来构成图像特征,能够很好地描述物体的形状和边缘,在图书图像中,不同图书图案的形状和边缘信息差别比较大,因此采用 HOG 特征描述效果较好。把每一个子图像分成 2×2 的子图像称为细胞单元 (cell), 把角度空间平均划分为 9 个 bin, 对每个细胞单元内所有像素的梯度值在各个 bin 区间进行直方图统计, 这样每个子图像就得到 36 维 HOG 特征。

3 联合特征相似性度量

3.1 子图像置信度

九宫格划分后,有些子图像的特征并不明显,如果与其他特征明显的子图像一起进行相似性度量反而会影响匹配的精度。因此对于每一幅拍摄图像,为它的子图像确定一个置信度,置信度高的子图像参与联合相似性度量,置信度低的子图像则不参与联合相似性度量。SIFT 特征关键点能够很好地表示局部区域特征的显著程度,一般来说 SIFT 点密集的区域图像特征显著也更有效,因此使用 SIFT 特征点密度作为子图像的置信度,选取特征点密度最高的 5 个子图像与库中图像的相应子图像进行相似性度量,然后计算联合相似度。

3.2 联合相似性计算

联合特征的相似性度量采用了“倒数比”的方式。计算方法为

$$c = \frac{2.0}{w_1 \times c_1 + w_2 \times c_2} \quad (5)$$

$$c_1 = n \times \sum_{i=1}^n \frac{1.0}{c_{1i}} \quad (6)$$

$$c_2 = n \times \sum_{i=1}^n \frac{1.0}{c_{2i}} \quad (7)$$

式中, c 表示联合距离; n 表示有效子图像数目,由人为指定; c_{1i} 表示第 i 个有效子图像的 SIFT 主方向角度直方图的距离; w_1 表示 SIFT 主方向角度直方图的距离的权重,由人为指定; c_{2i} 表示第 i 个有效子图像的 HOG 特征距离; w_2 表示 HOG 特征的距离的权重,由人为指定; c_1 为联合 SIFT 主方向角度直方图的距离; c_2 为联合 HOG 特征距离。

实验中 $n = 5, w_1 = 1.0, w_2 = 1.0$ 。

4 实验与分析

使用图书图像库数量为 50 000,是从互联网上下载的图书图像。使用的测试图像是用手机拍摄的真实图书图像。

由于在互联网上下载的图书图像分辨率不一致,并且有些图书图像有白色边框影响特征描述,因此首先去除白色边框提取图像的感兴趣区域(ROI),然后把 ROI 图像缩放到 500×500 ,然后按第 2 节的描述提取图像特征。

由于拍摄图像通常含有一部分背景,因此先把图书拍摄图像的边缘去除(上下左右各去除 5%),得到图书前景区域,然后把前景区域缩放到 500×500 与库图书图像一致,然后再提取特征检索。

部分实验结果如图 4 所示。



图 4 图书图像检索结果

Fig. 4 Retrieval results

图 4(a)(d)(e)的图书拍摄图有比较明显的高光,图 4(b)(d)的图书拍摄图有颜色和亮度变化。可以看出本文方法对高光、颜色和亮度变化有一定的鲁棒性。

图像库规模为 50 000 幅,在 3 G 内存、Windows XP SP3 系统 PC 机上,每一幅图像的检索时间在 $1 \sim 2$ s。

实验共测试了 100 幅图书拍摄图像,测试结果统计如表 1 所示。

表 1 检索结果统计

Table 1 Statistics of retrieval results

	第 1 匹配	前 3 匹配	前 3 无匹配
数目	66	78	22
比例	0.66	0.78	0.22

第 1 匹配表示第 1 个返回结果是正确的检索结果,检索精度为 66%;前 3 匹配表示前 3 个返回结果中含有正确检索结果,检索精度为 78%;前 3 无匹配表示前 3 个返回结果中无正确结果,认为检索错误。总体来说,检索精度和检索耗时在可接受的范围内。

对于检索错误的图像大体分为以下几类:

1)严重失真的拍摄图像。对于拍摄比较模糊、严重高光等造成严重失真的图像,特征点的提取会受到很大影响,从而导致特征匹配的不准确,影响检索结果。

2)图书拍摄不全的图像。由于图像中没有包含整本图书,因此基于九宫格的分块就把图书分割错误,每个子图像都会发生比较大的变化,因此会影响检索结果。

3)含有大量背景的图书图像。由于背景的大量存在,按照九宫格划分后有很多子图像会含有大量背景,从而影响图书图像的特征描述而造成检索错误。对于这个问题可以使用图像分割算法将图书分割出来然后进行检索,使用 GrabCut 算法^[8]把图书分割出来,然后把图书图像进行校正并缩放到 500×500 作为输入图像再按照本文方法进行检索,能解决一部分此类问题,但是由于图像分割问题一直是一个大问题,总是存在欠分割和过分割的问题,因此也会导致检索的精度下降。

4)图书旋转拍摄图像。由于九宫格的划分是固定的,所以对于有较大旋转的图书拍摄图像,子图像划分会出现大的偏差而导致检索错误。

5 结 论

提出一种基于九宫格的图书图像检索方法,把图书图像按照九宫格空间构建分割形成子图像,提取每个子图像的 SIFT 特征点密度和主方向角度直方图,结合 HOG 特征形成联合特征,然后按照“倒数比”的方式进行特征的相似性度量,检索的精度和耗时在可接受范围内。但是本文方法对拍摄图像有一定限制,如何消除对拍摄图像的限制和进一步提高精度是下一步需要继续研究的问题。

参考文献 (References)

- [1] Smeulders A W M, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12):1349-1380.
- [2] Lowe D G. Object recognition from local scale-invariant features [C]//Proceedings of International Conference on Computer Vision. Kerkyra, Corfu, Greece: IEEE Computer Society, 1999, 2:1150-1157.
- [3] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [4] Bay H, Tuytelaars T, Van G L. SURF: speeded up robust features[J]. Journal of Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [5] Yan K, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2004, 2:506-513.
- [6] Herve J, Matthijs D, Cordelia S. Improving bag-of-features for large scale image search[J]. International Journal of Computer Vision, 2010, 87(3):316-336.
- [7] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE Computer Society, 2005, 1:886-893.
- [8] Rother C, Kolmogorov V, Blake A. Grabcut: interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics, 2004, 23(3):309-314.