

基于局部对称性的特征点加工策略及应用

郭一超 李清勇 孙靳睿 黄雅平 田 媚

(北京交通大学计算机与信息技术学院 北京 100044)

摘要 在基于词袋模型的图像检索框架中,图像包含的 SIFT 特征点往往数量比较大,特征不够强。因此图像检索系统的效率和性能往往受影响。基于 SIFT 特征点的性质和视觉显著性原理,提出了 SIFT 特征点的局部对称性度量方法,并且在图像检索框架中嵌入了基于对称性的 SIFT 特征点过滤方法和加权策略,以提升 SIFT 特征点的利用效率。在牛津大学建筑物图像集上的实验结果表明,提出的基于对称性的 SIFT 特征点选择策略能有效地提高图像检索的性能。

关键词 对称性,图像检索,显著性,SIFT 特征点

中图分类号 TP391.3 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.11.057

Feature Points Processing Strategies Based on Local Symmetry and its Application

GUO Yi-chao LI Qing-yong SUN Jin-rui HUANG Ya-ping TIAN Mei

(Department of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

Abstract In the image retrieval framework based on Bag of Words (BoW) model, images usually contain a large number of SIFT feature points whose features are not strong enough and have influence on the efficiency and performance of the image retrieval system. Based on properties of SIFT feature points and principle of visual saliency, this paper proposed a local symmetry measure method for SIFT feature points, and embedded two symmetry processing strategies in the BoW image retrieval framework: filtering method and weighting strategy. The experimental results on the Oxford Buildings dataset show that the selection strategy of SIFT feature points based on symmetry can effectively improve the performance of image retrieval systems.

Keywords Symmetry, Image retrieval, Saliency, SIFT feature points

1 引言

词袋模型(Bag of Words, BoW)是当前图像检索的最重要方法之一。该方法首先抽取图像的局部特征点以及特征点的描述子,然后基于聚类生成的视觉词典(Visual Codebook)量化特征描述子,即把特征描述子映射到特定的视觉单词,最后应用倒排索引等技术实现图像检索。在 BoW 框架中,特征点抽取是最基础和最重要的步骤,SIFT 特征点是应用最为广泛的方法之一^[1]。SIFT 算法是一种基于尺度空间的,对图像缩放、旋转和部分仿射变换保持不变性的图像局部特征提取算法,Mikolajczyk 等人针对包括 SIFT 算子在内的 10 种著名局部特征点检测算法进行比较,验证了 SIFT 及其扩展算法是最健壮和有效的^[2]。但是发现在基于 BoW 的图像检索中,SIFT 描述子也存在以下不足:1)图像的局部特征点是通过在 DoG(Difference of Gaussians)尺度空间上探测极值点来获得的,部分极值点是不稳定的噪声点;2)大部分图像生成的 SIFT 特征点数量巨大,过多的 SIFT 特征点将造成索引空间增大、搜索时间延长和检索准确度下降等后果。

针对 SIFT 描述子的上述不足,Lowe 等提出了两条删除局部特征点的规则:过滤对比度小的特征点和忽略边缘响应度强的特征点。该方法在一定程度上能消除不稳定的局部极值点,但是我们的实验结果表明,过滤后的 SIFT 特征点规模仍然比较大,比如在牛津大学建筑数据集^[3]上每幅图像平均有近 1000 个特征点。Li 等人采用数据驱动的监督学习方法来选取稳定的局部特征点,并提出了改进后的基于排序的 SIFT 算法(Rank-SIFT)^[4]。该方法能够计算每一个特征点的稳定度,还可以根据实际应用中的效果和效率等因素来控制被保留特征点的数量,但是该方法需要额外的训练数据等先验信息,增加了对系统和用户的要求。Zhang 等人提出了间接的改进方法,他们不考虑单个的局部特征点,而是将局部特征点以组为单位进行相关的处理,通过局部特征组引入了特征点之间的空间信息^[5]。本文提出了一种基于局部对称性的 SIFT 特征点选择策略,该策略在检测 SIFT 特征点后,计算 SIFT 特征点局部区域的对称性,然后基于局部对称性来选择 SIFT 特征点。局部对称性是一种数据驱动的、自底向上的显著性度量方法,对称性感知也是生物视觉系统早期视

到稿日期:2013-06-25 返修日期:2013-08-16 本文受国家自然科学基金项目(61272354,61273364,61105119),中央高校基本科研业务费(2011JBZ005,2012JBM039,2012JBM027),北京交通大学轨道交通控制与安全国家重点实验室自主研究课题(RCS2012ZT007)资助。

郭一超(1989—),女,硕士生,主要研究领域为模式识别与机器学习,E-mail:dorothyguo@163.com;李清勇(1979—),男,博士,副教授,主要研究领域为视觉智能与数据挖掘;孙靳睿(1990—),男,硕士生,主要研究领域为信息检索和图像处理;黄雅平(1974—),女,博士,教授,主要研究领域为深度学习和模式识别;田媚(1980—),女,博士,讲师,主要研究领域为视觉注意机制。

觉过程的重要特征^[6]。与 Lowe 等的方法^[1]相比,本文提出的策略选择目标有更强的针对性,也更加符合人类视觉注意机制;另外,本文方法是一种无监督的、视觉注意相关的处理过程,这明显不同于 Rank-SIFT 方法,其在图像检索等应用中实现更加方便。本文的主要贡献包括以下两方面:

1)把基于注意机制的显著性测度引入 SIFT 特征点选择,揭示了 SIFT 特征点的局部对称性特征,提出了基于局部对称性的 SIFT 特征点过滤和加权两种策略;

2)在 BoW 图像检索框架中,集成了 SIFT 特征点选择策略,并在牛津建筑物数据集上验证了该策略的有效性。

本文第 2 节回顾了显著性模型和图像检索的相关工作;第 3 节阐述了 SIFT 特征点的局部对称性特征和选择策略;第 4 节给出了本文的实验结果;最后对本文研究进行了总结和展望。

2 相关工作

视觉注意计算模型一般使用两种信息引导注意力的转移:自底向上基于图像显著性的信息和自顶向下基于任务的信息。本节深入分析了视觉显著性及其计算模型的相关内容,并进行了基于 BoW 的图像检索的研究,尝试将视觉显著性应用在图像检索上。

2.1 视觉显著性计算模型

Treisman 和 Gelade 所提出的“特征整合理论”可以认为是最早的注意模型,该理论阐述了哪些视觉特征是重要的,以及如何组合这些特征来引导人们的视觉注意^[7]。Itti 等人首次实现并验证了上述理论,提出了有效的可计算视觉注意模型。该模型分别在亮度、颜色和方向 3 个不同的特征通道上建立高斯金字塔,在图像金字塔上应用中央周边差(Center Surround Difference)操作生成特征显著图^[8]。Itti 模型极大地促进了可计算显著性模型的发展,各种各样基于不同注意机制假设的模型涌现出来,Borji 和 Itti 对当前基于显著性的注意模型进行了分类和对比研究^[9]。当前主流的视觉显著性计算模型可以归纳为以下 4 种类型。

1)基于认知机理的显著性模型。该类模型直接或间接地受认知概念和神经机理的启发,实现对认知过程的功能模拟。Itti 模型是一种典型的基于认知机理的显著性模型^[8]。对称性是一类丰富的视觉特征,不仅是人造物体和建筑,大多数的自然生物也有着高度的对称性。很多研究表明人类对于对称性的感知是非常强烈的,并且该感知过程工作在前注意阶段,是早期视觉加工的重要内容^[10]。Kootstra 提出了一种基于局部对称性的注意模型,与基于对比度的显著性模型相比,该模型具有更好的集中性^[6]。Le Meur 等人提出了基于人类视觉系统(HVS)结构的自底向上显著性方法,在这个模型中实现了对比敏感度函数、感知视觉掩蔽和中央周边交互等特征^[11]。

2)基于统计原理的显著性模型。Li 等人提出一个视频中视觉注意的贝叶斯多任务学习框架,亦即将多尺度小波分解与多任务学习算法训练出的不同的自顶向下组件互相结合来指导自底向上的显著性信息。该模型也能够学习不同的策略将自底向上和自顶向下的图相结合来获得最终的注意图^[12]。Pang 等人提出了一个基于信号探测理论的视觉注意随机模型,由于人类视觉注意没有确定性,同一时间对于相同

的视觉输入,不同的人可能注意不同的位置,因此他们根据动态贝叶斯网络来预测在一个视频场景中人类通常会专注于哪里^[13]。

3)基于机器学习的显著性模型。该类模型根据记录下来的眼睛注视点或是有标签的显著区域进行学习并获取显著性模型。值得注意的是,因为其会用到引导自顶向下注意的特征,这类模型可能不仅仅是自底向上的。Judd 等人根据人眼注视数据训练了一个线性 SVM,用一组低级、中级和高级图像特征来定义显著位置,并且在注视位置和随机位置上的特征向量类标签分别被标记为+1 和-1,其实验结果表明联合以上所述的所有特征以及到图像中心的距离会产生最佳人眼注视点预测结果^[14]。Gao 等人用判别式显著模型进行视觉识别,该模型在 PASCAL 2006 数据集上表现出良好的性能^[15]。

4)基于非空间域的显著性模型。该类模型应用频域分析方法从图像的频域中导出图像的显著图,虽然在生物学上的合理性不是非常明确,但是实现容易、效果比较好。针对生物视觉系统能略过冗余的(频繁出现的、信息量少的)特征、却对稀少的(信息量丰富的)特征做出响应的现象,Bian 和 Zhang 等提出了谱白化模型,取得了较好的效果^[16]。

2.2 基于 BoW 的图像检索

BoW(Bag of Words)理论最早起源于自然语言处理技术,是信息检索领域中常用的文档表示方法。该理论将一个文档看作是若干词汇的集合,文档中每个单词的出现都认为是独立的^[17]。牛津大学的 Sivic 和 Zisserman 最早提出了基于 BoW 的图像检索方法^[18]。该方法首先抽取图像的局部特征点以及特征点的描述子,然后基于聚类生成的视觉词典量化特征描述子,即把特征描述子映射到特定的视觉单词,最后应用倒排索引等技术实现图像检索。

基于 BoW 的图像检索框图如图 1 所示,它包括后台管理子系统和前台检索子系统。后台子系统首先抽取图像库中每幅图像的 SIFT 描述子,并且把每一个描述子映射到特定的视觉单词,然后统计视觉单词在每幅图像和图像库中的频次,并依据特定的加权策略(比如 TF/IDF 策略)生成倒排文件索引。需要注意的是,在 SIFT 描述子量化前,需要聚类生成视觉词典。前台子系统包括抽取查询图像的 SIFT 描述子、视觉单词量化和相似度匹配等过程。

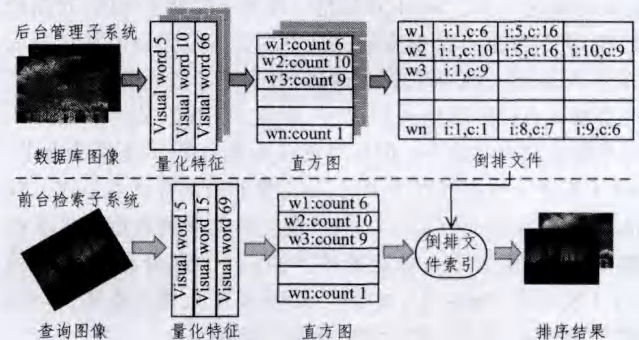


图 1 基于 BoW 模型的图像检索流程,倒排文件的每一行记录 (i,c) 表示视觉单词 w 在图像 i 中出现的次数为 c

在上述过程中,SIFT 特征点是非常关键的要素,特征点的稳定性会影响检索的性能,特征点的数量会影响检索的效率。下一节将介绍基于局部对称性的 SIFT 特征点选择策略。

3 基于对称性的 SIFT 特征点选择及图像检索

对称性作为一种丰富的视觉特征,由于在视觉注意中的重要地位,已经被成功地应用在大量的研究中。本节基于已有的研究结论更加深入地探究了局部特征点的对称性,并将其运用在图像检索当中,以期达到优化检索结果的目的。

3.1 对称性算子

给定任意点 P , 其对称核(Symmetry Kernel)定义为以 P 为中心的对称区域, 记为 $SK(P)$, 图 2 展示了一个 7×7 的方形对称核。如果给定 $SK(P)$ 中以 P 中心对称的点 P_i 和 P_j , 其对称性定义为 $s(i, j)$, 那么综合 $SK(P)$ 中所有对称点对的对称性值即可得到点 P 的对称性。

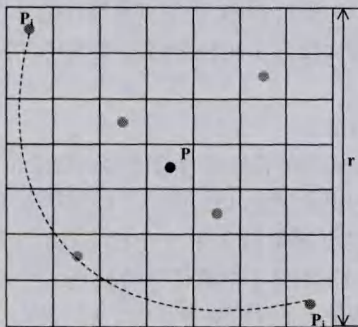


图 2 点 P 的对称核和对称点对示意图

点对 P_i 和 P_j 的对称性基于点对的灰度梯度来计算。应用 Sobel 算子, 可得到点 P_i 处的梯度向量 $Grad_i = (g_i^x, g_i^y)$, 其中 g_i^x 和 g_i^y 分别表示点 P_i 在水平和垂直方向的 Sobel 算子响应值。 $Grad_i$ 的模 m_i 和 θ_i 方向由下式决定:

$$m_i = \sqrt{(g_i^x)^2 + (g_i^y)^2}, \theta_i = \tan^{-1}(g_i^y / g_i^x)$$

P_i 和 P_j 的方向对称性 $c(i, j)$ 定义如下:

$$c(i, j) = (1 - \cos(\gamma_i + \gamma_j)) \cdot (1 - \cos(\gamma_i - \gamma_j)) \quad (1)$$

其中, γ_i 表示点 P_i 的梯度方向与直线 $\overrightarrow{P_i P_j}$ 的夹角, 如图 3 所示, 令 α 表示直线 $\overrightarrow{P_i P_j}$ 与水平方向的夹角, 则有 $\gamma_i = \theta_i - \alpha$ 。当 $\gamma_i + \gamma_j = \pi$, 也就是像素点对的梯度方向关于对称轴 a 镜面对称时, 式(1)中的第一项能取最大值, $c(i, j)$ 的值将比较高; 同样的道理, 当像素点对的梯度方向完全相反或者说关于点 P 对称, 即 $\gamma_i - \gamma_j = \pi$ 时, $c(i, j)$ 的值也将比较高。但是当 $\gamma_i - \gamma_j = 0$, 即像素点对的梯度方向相同时, 式(1)中的第二项取值为 0, $c(i, j)$ 的值将比较低。

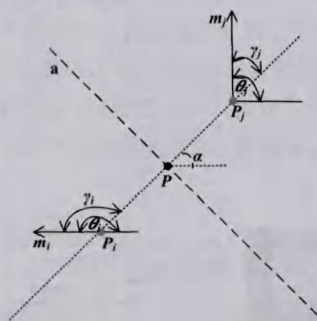


图 3 利用一个像素点对的灰度梯度来计算该像素点对的贡献

考虑离中心点 P 的距离和梯度大小, 点对 P_i 和 P_j 的加权对称性 $s(i, j)$ 定义如下:

$$s(i, j) = d(i, j, \sigma) \cdot c(i, j) \cdot \log(1 + m_i) \cdot \log(1 + m_j) \quad (2)$$

其中, $d(i, j, \sigma)$ 是标准偏差为 σ , P_i 和 P_j 之间距离的高斯加权函数。梯度相关项确保了只有明显的边缘对局部对称值起作用, 另外对数运算减弱了大梯度值的影响。

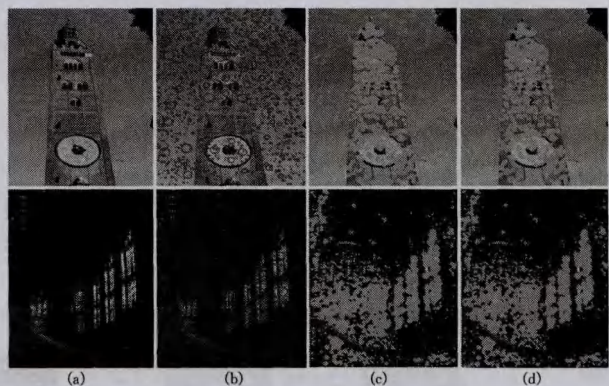
计算对称核 $SK(P)$ 中所有对称像素点对的加权对称值, 则可得到点 P 的对称值 $S(P)$:

$$S(P) = \sum_{(i, j) \in SK(P)} s(i, j) \quad (3)$$

3.2 SIFT 特征点的局部对称性度量

SIFT 特征点本质上是高斯拉普拉斯空间中的极值点, 在图像空间域上, SIFT 特征点对应图像中与周围存在较大颜色或者灰度差异的局部区域^[1]。与 Harris 角点等特征点相比, SIFT 特征点的稳定性好、抗噪声能力强。但是 SIFT 特征点数目比较多, 需要比较多的存储和计算资源; 从生物视觉机理分析, SIFT 点不见得是显著性比较强、生物视觉系统关注的点。SIFT 特征点的对称性度量和基于对称性的选择策略可以改善上述问题。

SIFT 特征点的对称性度量包括两个步骤: 1) 抽取图像的 SIFT 特征点, 记为 $\{P_1, P_2, \dots, P_n\}$; 2) 计算每一个特征点的对称性值 $S(P_i)$ 。图 4 展示了图像的 SIFT 点和对称性属性: 图 4(a) 是输入图像; 图 4(b) 显示图像的 SIFT 特征点, 每一个特征点对应一个以 SIFT 点坐标为圆心、尺度大小为半径的空心圆; 每个 SIFT 点的对称性用图 4(c) 中的实心圆表征, 实心圆的圆心为 SIFT 特征点, 半径等于该点对应的局部对称性值的 0.2 倍。在提取 SIFT 特征点时, 我们采用了 Lowe 原文建议的规则过滤了不稳定的 SIFT 点。从图 4(b) 可以看出, 无论是背景单一(第一行图像)还是背景复杂(第二行图像)的图像, 虽然不少 SIFT 点落在拐角或边缘处, 非常具有代表性, 但是 SIFT 特征点都比较多, 即使特征不明显的背景区域也包含大量的 SIFT 点。在图 4(c) 中, 背景处的 SIFT 特征点的对称性值一般都比较低, 而前景 SIFT 点的对称性值保持一个比较高的水平。图 4(d) 显示了过滤对称性值小于 10 的 SIFT 点。不难看出, 过滤后的 SIFT 点保留了图像的绝大部分特征明显的前景点。综上所述, 对称性可以为 SIFT 点的视觉显著性或者重要性提供一种有效的度量方法。



从左往右依次为输入图像、SIFT 点示意图、对称性示意图和过滤(阈值等于 10)后的对称性图

图 4 SIFT 特征点的对称性度量实例

统计分析 SIFT 特征点的对称性值的概率分布, 我们发现低对称性的 SIFT 点在单个图像中往往出现的概率比较

大。图 5(a)和图 5(b)分别展示了图 4 中两幅典型图像的 SIFT 特征点的对称性的直方图,横坐标为 SIFT 点的对称性值,纵坐标为局部对称性值落在某个范围内的 SIFT 特征点的个数。图 4 中的两幅图像具有很大的差别:上图背景相对简单,前景也仅有一栋建筑,看上去一目了然;下图前景和背景都含有非常丰富的信息,尤其是背景非常复杂且涵盖信息量大。尽管这两幅图像有着很大的差别,但分析其 SIFT 特征点集的对称性统计直方图后可发现:无论图像的属性如何,它们都存在很多局部对称性小的、不稳定的 SIFT 特征点。由于 SIFT 特征点的多量性,即使含有少量物体且背景简洁的图像也会产生大量的 SIFT 特征点。例如图 4 中天空背景产生了大量的 SIFT 特征点,占有 SIFT 点总数的一半左右,这些 SIFT 特征点不仅没有丰富的视觉信息,在实际的应用中(如图像检索)也没有很大的利用价值,如果不进行有效的选择,只会增加应用处理的负担。

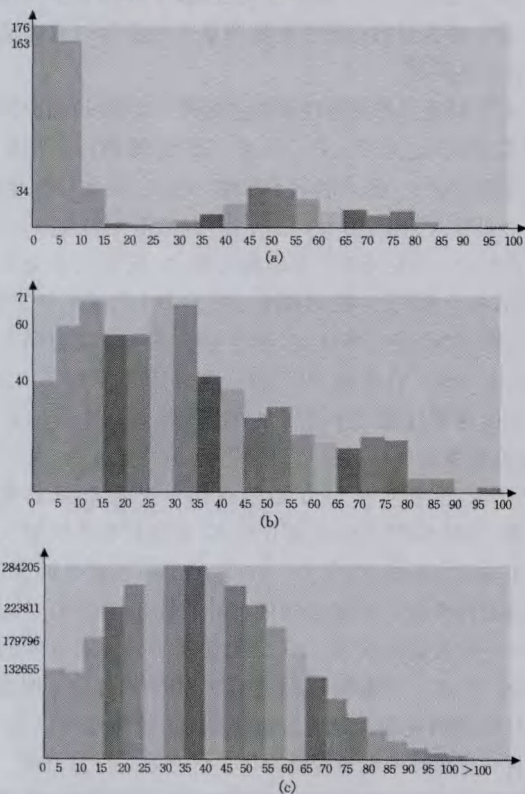


图 5 两幅典型图像和整个牛津建筑物数据集的 SIFT 特征点局部对称性统计直方图

进一步统计分析一个图像集中 SIFT 特征点的对称性值分布,我们发现低对称性的 SIFT 点所占的比例也不低。我们对牛津建筑物数据集(The Oxford Buildings Dataset)^[3]中所有的 5063 幅图像提取 SIFT 特征点,并计算每个 SIFT 特征点的局部对称性,其直方图如图 5(c)所示。尽管在 SIFT 特征点抽取过程中把输入图像的最大边限制在 370 以内,并

且根据 Lowe 提出的两种手工规则进行了前期筛选,我们仍然得到了将近 327 万个 SIFT 特征点。在这些点中,大约 20% 的 SIFT 特征点具有较低的局部对称性,这些 SIFT 特征点都较不稳定,极有可能是具有干扰性或仅有少量视觉信息的噪声点。

3.3 基于局部对称性的图像检索

本文改进了传统的 128 维 SIFT 特征向量的表示,在生成 BoW 的过程中引入了局部对称性这一丰富的视觉信息。首先将图像库中的每幅图像都做如下相同处理:提取图像中的 SIFT 特征点,根据每个特征点的坐标计算该点的局部对称性,并将其保存,作为 SIFT 特征向量的一维。一个 SIFT 特征点的局部对称性越高,代表这个 SIFT 特征点越重要。局部对称性作为衡量 SIFT 特征点重要性的一个标准。本文根据局部对称性设计了两种策略加工 SIFT 特征点:过滤和加权。

1) 过滤方法

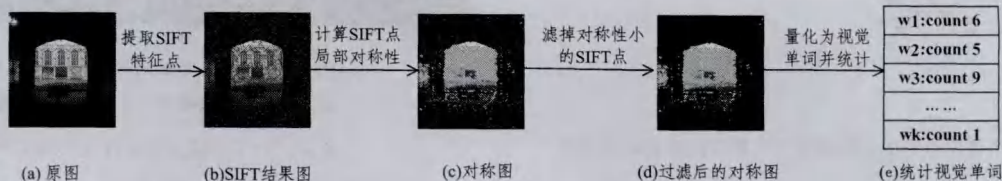
局部对称性高的特征点含有丰富的视觉信息,能够明显地引起人眼的视觉注意,稳定度高,反之亦然。假设任意图像 I 通过 SIFT 算法抽取含有 m 个特征点的序列 $\{x_i \in I, i=1, 2, \dots, m\}$,对于图像 I 中的任意特征点 $x_i \in I$,稳定的特征点定义为该特征点的局部对称性大于某个阈值。图像 I 中稳定的特征点数目定义如下:

$$N(x_i \in I) = \sum_m R(x_i) \quad (4)$$

其中, $R(x_i)$ 定义为:

$$R(x_i) = \begin{cases} 1, & S(x_i) > threshold \\ 0, & \text{否则} \end{cases} \quad (5)$$

$S(x_i)$ 表示特征点 x_i 局部对称性的大小, $threshold$ 为局部对称性阈值。如图 6(b)所示,传统的 SIFT 算法可能在一幅图像上生成大量的局部特征点,我们希望这些特征点尽量出现在信息含量比较丰富的建筑物或物体上,而尽可能少地出现在背景中。图像局部特征点周围小区域上的对称性能够描述图像部分内容信息,并且在局部和全局扰动变化中具有稳定性,因此可以通过合理地选择对称性阈值来尽可能多地保留那些具有高稳定度的局部特征点,去掉稳定度低的特征点。显然图 6(d)中根据本文所提方法过滤后的局部特征点质量要优于未经过滤的图 6(c)。图像检索系统需要对图像中的特征点进行聚类、构建词表和进行特征量化等多个处理步骤,因而局部特征点的稳定性和质量好坏直接影响到后续算法处理的效果。如图 6 所示,本文在每个 SIFT 特征点量化为视觉单词之前,通过选择合适的对称性阈值过滤局部对称性小的、不稳定的或视觉信息弱的 SIFT 特征点,使得图像中的噪声点尽可能地占最小的比重,而真正有意义的特征点能够占绝大部分的比重,以期达到提高检索效率和准确率的目的。



图(e)的每行记录表示第 i 个视觉单词出现的次数

图 6 根据局部对称性过滤 SIFT 特征点

2) 加权策略

假设单词表 V 中含有 k 个视觉单词 $\{w_j \in V, j=1, 2, \dots, k\}$, 任意图像 I 通过 SIFT 算法抽取含有 m 个特征点的序列 $\{x_i \in I, i=1, 2, \dots, m\}$, 则对于任意的视觉单词 $w_j \in V$, 根据局部对称性进行加权后其在图像中出现的次数 $C(w_j)$ 定义如下:

$$C(w_j) = \sum_m M(x_i) \quad (6)$$

其中, $M(x_i)$ 定义为:

$$M(x_i) = \begin{cases} S(x_i), & x_i \text{ 量化为单词 } w_j \\ 0, & \text{否则} \end{cases} \quad (7)$$

$S(x_i)$ 表示特征点 x_i 局部对称性的值。未加权之前, 若一个特征点量化为某个视觉单词, 则该视觉单词在图像中的出现次数加 1。现采用局部对称性进行加权, 若某个 SIFT 特征点的对称性越高, 则其稳定性越好, 对应的视觉单词权重越大, 如式(7)所示, 根据该视觉单词在图像中的出现次数增加较大的权值。如图 7 粗线箭头所示, 特征点 1 和特征点 i 都量化为视觉单词 1, 未加权之前, 视觉单词 1 在图像中的出现次数为 2, 加权后视觉单词 1 在图像中的出现次数为 $S_1 + S_i$ 。本文是通过使用 TF/IDF 加权打分机制的图像检索实验来评价所提方法有效性的, 由于在图像检索时根据局部对称性加权改变了图像中视觉单词的出现次数, 因此改变了视觉单词的词频(TF/IDF 中的 TF 项)。

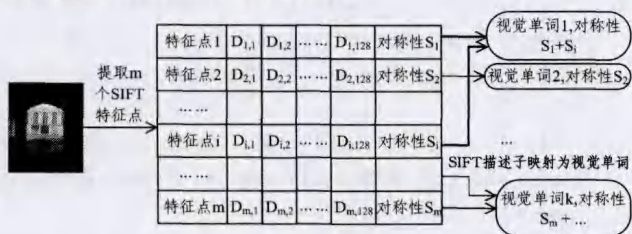


图 7 引入局部对称性的 SIFT 特征向量表示及 BoW 的生成, $D_{i,j}$ 表示第 i 个 SIFT 特征点向量表示的第 j 维

4 实验

我们在牛津建筑物数据集(The Oxford Buildings Dataset)^[3]上进行图像检索实验来进一步评估本文所提方法。该数据集由牛津城区 11 个建筑物为主的 5063 张图像和 55 张查询图像组成, 即每个建筑物有 5 张查询图像。实验目的是比较传统 SIFT 算法和本文改进后的 SIFT 算法应用于图像检索中的效果。在基于 BoW 的图像检索框架中引入 3.3 节所述方法进行实验, 对于每张查询图像, 可得到整个数据集中与该图像最相关的图像排序列表, 将其与基准作比较, 计算出均值准确率。均值准确率是通过准确率和召回率在二维坐标上围出的面积来定义的。最终, 我们将 55 幅查询图像的平均均值准确率(mean Average Precision, mAP)作为检索性能指标, 显然 mAP 值越高表明算法越有效。需要具体指出的是, 本实验中通过 K-Means 方法聚类产生视觉单词表。

4.1 过滤方法实验结果

我们比较了传统 SIFT 算法和引入局部对称性过滤策略的 SIFT 算法, 实验选取 4 个局部对称性阈值参数, 视觉词表中的单词数均为 10000, 图像检索结果如表 1 所列。

表 1 通过阈值过滤 SIFT 特征点的检索实验结果比较

局部对称性阈值	0	10	15	20
mAP(%)	35.5	36.66	35.94	36.45

从表 1 结果来看, 相比传统 SIFT 算法(局部对称性阈值为 0), 本文提出的过滤方法在其他非零对称性阈值参数设置情况下, 明显提高了图像检索的正确率, 取得了更好的效果。分析表 1 可知, 对称性阈值和检索实验结果 mAP 之间并没有线性关系, 阈值参数为 15 的情况下其检索效果略差于参数为 10 和 20 的情况, 且在参数取值为 10 时效果最佳。对于不同的图像集最佳的阈值设置可能也不相同, 这与图像集中图像的具体属性和特点有关。因此在实际应用时, 需要多次尝试不同的阈值参数平衡被保留特征点质量与数目之间的关系, 以期达到最好的检索效果。

4.2 加权策略实验结果

我们在 3 种不同的视觉单词数目设置情况下, 分别比较了传统 SIFT 算法和引入局部对称性加权策略的 SIFT 算法在图像检索实验中的效果, 实验对比结果如表 2 所列。

表 2 用局部对称性加权前后的检索实验结果比较(mAP: %)

视觉单词数	未加权	对称性加权
10000	35.5	37.45
20000	38.4	42.99
50000	46.4	48.51

从表 2 结果来看, 在不同的视觉单词数目设置情况下, 引入局部对称性加权策略的 SIFT 算法在图像检索正确率方面都明显优于传统的 SIFT 算法。无论是传统的 SIFT 算法还是本文改进后的方法, 随着视觉单词数的增加, 检索实验结果 mAP 随之增长, 两者成正比的关系。数量巨大且丰富的视觉单词显然能够更好地描述特征点, 从而提高检索结果。但是如果视觉单词的数目过大, 就会在特征点量化和匹配过程中耗费大量的时间, 严重降低检索效率, 这是在实时检索系统中无法容忍的。

4.3 图像检索结果示例

在牛津建筑物数据集上做图像检索实验的部分结果示例如图 8 所示, 图中框中的图像为查询图像, 其余的图像是对应查询图像的返回结果; (a) 行和 (b) 行检索正确率最高; (c) 行检索正确率中等; (d) 行和 (e) 行检索正确率较差, 存在错误的检索结果。



第 1 列为查询图像, 从上往下检索正确率依次降低

图 8 牛津建筑物数据集检索结果示例

由图 8 可知, 与传统的 SIFT 算法相比, 本文所提方法能够在图像检索实验上取得较好的效果, 但是仍然存在一些不足有待改进。本文提出的两种评估 SIFT 特征点的方法虽然能够达到过滤掉一部分噪声点或使稳定度高的特征点占更大

比重的目的,但仍然会遇到如下问题:有的图像内容非常丰富,导致其提取到的 SIFT 特征点数量比一般图像要大得多,即使引入过滤或加权策略后,仍会保留很多特征点。无论对于什么样的查询图像,这类图像都能够以“符合查询”的结果返回而造成误匹配,如图 8(e)行第 4 幅图像就属于这样的情况。

结束语 本文主要探究了如何使用非监督学习的方法来优化局部特征点检测子算法的问题,首次提出根据局部对称性来选择 SIFT 特征点的过程,并通过图像检索实验验证了本文所提方法优于传统的 SIFT 算法。我们所提出的基于局部对称性的 SIFT 特征点评估方法具有可扩展性,可以扩展到优化其他的检测子算法,比如 Harris 检测子等。

近些年来,对 SIFT 算法的研究工作不断涌现,然而对如何从候选的 SIFT 特征点中选择出稳定的局部特征点的研究却很少。由于对称性对人类视觉注意的重要影响,我们提出了基于局部对称性评估 SIFT 特征点稳定性的方法。我们下一步的工作包括以下 3 方面:1)将基于局部对称性评估 SIFT 特征点的方法扩展到 Harris 检测子等其他算法上,也可以借鉴其他检测子的优点进一步改进现有的评估方法;2)已有的文献已经提出了大量的对称性算子,接下来的工作可以尝试用不同的算子来计算对称性,从中选择出最优算子或将一些算子联合来进行实验,从而提高对 SIFT 特征点评估的准确性;3)在实际应用中,需要设计一种策略来权衡被保留 SIFT 特征点的数目,该策略既要保证被保留的 SIFT 特征点具有较高的稳定性,又要保证其数目尽可能小,以降低后续处理的负担。

参 考 文 献

[1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2):91-110

[2] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10):1615-1630

[3] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2007:1-8

[4] Li B, Xiao R, Li Z, et al. Rank-SIFT: Learning to rank repeatable local interest points[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011:1737-1744

[5] Zhang S, Huang Q, Hua G, et al. Building contextual visual vo-

cabulary for large-scale image applications[C]// Proceedings of the international conference on Multimedia. 2010:501-510

[6] Kootstra G, de Boer B, Schomaker L R. Predicting eye fixations on complex visual stimuli using local symmetry[J]. Cognitive computation, 2011, 3(1):223-240

[7] Treisman A M, Gelade G. A feature-integration theory of attention[J]. Cognitive psychology, 1980, 12(1):97-136

[8] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259

[9] Borji A, Itti L. State-of-the-art in visual attention modeling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1):185-207

[10] Wagemans J. Parallel visual processes in symmetry perception: Normality and pathology [J]. Documenta ophthalmologica, 1998, 95(3/4):359-370

[11] Le Meur O, Le Callet P, Barba D, et al. A coherent computational approach to model bottom-up visual attention[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5):802-817

[12] Li J, Tian Y, Huang T, et al. Probabilistic multi-task learning for visual saliency estimation in video[J]. International journal of computer vision, 2010, 90(2):150-165

[13] Pang D, Kimura A, Takeuchi T, et al. A stochastic model of selective visual attention with a dynamic Bayesian network[C]// IEEE International Conference on Multimedia and Expo. 2008:1073-1076

[14] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look[C]// IEEE 12th International Conference on Computer Vision. 2009:2106-2113

[15] Gao D, Han S, Vasconcelos N. Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(6):989-1005

[16] Bian P, Zhang L. Biological plausibility of spectral domain approach for spatiotemporal visual saliency[J]. Advances in Neuro-Information Processing, Springer, 2009, 5506:251-258

[17] Csurka G, Dance C, Fan L, et al. Visual categorization with bags of keypoints[C]// ECCV Workshop on Statistical Learning in Computer Vision. 2004:22

[18] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos[C]// Proceedings of Ninth IEEE International Conference on Computer Vision. 2003:1470-1477

(上接第 285 页)

[7] Krogager E. A new decomposition of the radar target scattering matrix [J]. Electronic Letter, 1990, 26(18):1525-1526

[8] Cloude S R, Pottier E. A Review of target decomposition theorems in radar polarimetry [J]. IEEE Transactions on Geoscience and Remote Sensing, 1996, 34(2):498-518

[9] Lee J S, Grunes M R, Pottier E, et al. Unsupervised terrain classification preserving polarimetric scattering characteristics [J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(4):722-731

[10] Lee J S, Grunes M R, Ainsworth T L, et al. Unsupervised classification using polarimetric decomposition and complex Wishart

classifier [J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(5):2249-2348

[11] Li S Z. Markov random field modeling in image analysis(3rd Edition)[M]. Springer-Verlag London, 2009

[12] Wu Yong-hui, Ji Kang-feng, Yu Wen-xian, et al. Region-based classification of polarimetric SAR images using Wishart MRF [J]. IEEE Geoscience and Remote Sensing Letters, 2008, 5(4):668-672

[13] Lee J S. Improved sigma filter for speckle filtering of SAR imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(1):202-213