

一种适用于智能手机的图像识别算法

桂振文¹, 刘 越², 陈 靖², 王涌天^{1,2}, 徐志伟²

(1.北京理工大学计算机学院,北京 100081;2.北京理工大学光电学院北京市混合现实与新型显示工程技术研究中心,北京 100081)

摘 要: 针对目前常用的图像识别算法运算复杂和内存占用量大,不能很好的应用于移动平台等问题,本文提出了一种适用于智能手机的图像识别算法:首先,通过使用 BRISK 特征点检测算法提取图像特征和低字节的 FREAK 描述符对特征进行表述,解决了特征检测时间长和特征描述符内存占用大的问题;其次,将智能手机的重力信息添加到图像特征中改善了 BRISK 特征的区分能力,解决了相似结构特征难以区分的问题;最后,建立描述符的多级索引,实现相似描述符的快速查找,解决了描述符匹配问题.实验结果表明,本文提出的算法能有效地运行在资源受限的智能手机上实现对场景的实时识别.

关键词: 图像识别; FREAK 算法; BRISK 算法; 智能手机

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2014)08-1487-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2014.08.005

A Novel Image Recognition Algorithm for Smartphones

GUI Zhen-wen¹, LIU Yue², CHEN Jing², WANG Yong-tian^{1,2}, XU Zhi-wei²

(1. School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China;

2. Beijing Engineering Research Center for Mixed Reality and Advanced Display, School of Optoelectronics, Beijing Institute of Technology, Beijing 100081, China)

Abstract: At present, the conventional image recognition algorithm cannot be well applied to mobile platforms due to their large computational cost and high storage requirements. In this paper, a novel image recognition algorithm for smartphone is proposed: Firstly, BRISK detection algorithm and low-byte FREAK descriptor are used to extract image feature and represent image respectively, which solves the problems of long detection time and large memory footprint; Secondly, BRISK features are improved to accomplish the distinction between the similar structure features by adding the gravity information of smartphone to the image feature; Finally, multi-level index of descriptors have been established to achieve fast finding on similar descriptors which solve the matching problem. The experimental results show that the proposed algorithm can be effectively run on the resource-constrained general smartphone to achieve scene recognition in real-time.

Key words: image recognition; freak algorithm; frisk algorithm; smartphone

1 引言

图像识别是计算机视觉领域中最活跃的研究课题之一,主要研究目标在于利用计算机的强大运算能力,帮助人类自动处理海量物理信息,识别各种不同模式的目标代替人的部分脑力劳动.图像识别综合了很多学科内容,其中包括计算机科学与技术、物理学、统计学以及神经生物学等,广泛应用于地质勘探、图像遥感、机器人视觉、生物医学等多个领域.图像识别技术在个人计算机以及嵌入式终端设备上已有很多成熟的应用案例,随着手机功能的不断增强,该技术也逐渐被应用到智能手机.但智能手机本身的弱处理能力和低内存的局限性

对计算复杂的识别算法提出了考验.

图像识别的基本原理是:首先对目标图像进行特征点检测,然后再对特征点进行描述,最后通过特征点匹配来确定目标图像对应样本库中最相似的图像.关于智能手机上大规模图像识别目前有两种研究思路:一种思路是依靠客户-服务器(C/S)模型,只在智能终端获取图像和显示结果,而在 PC 端完成特征点提取、匹配等计算量大的复杂算法,通过无线通信进行交互,这种方法易受到无线网络的影响,在网络拥塞的情况下效果较差.如文献[1]通过在智能终端获取带 GPS 信息的图像发送到服务器端进行识别.文献[2,3]在 PC 端采用分层单树树(Hierarchical-VOC)的算法实现大规模图像的快

速识别.文献[4,5]通过采用机器学习 SVM 的方法实现大规模图像的高准确率的识别.另一种思路对复杂算法进行优化和改进,在智能手机上完成图像识别算法.如文献[6]通过降低的图像特征提取、特征描述和匹配的计算复杂性,实现了智能手机端上的小规模图像识别.文献[7]通过 Hash 函数将高维的特征向量进行的压缩,用 30bit 的汉明码来表示高维特征,用汉明距离来度量描述符相似,仅用一条汉明指令就能很快进行特征点匹配.

对于特征点检测,研究者提出了大量的特征点检测算法.例如:Schmid 等^[8]提出 Harris-Laplace 算法,将 Harris 和高斯拉普拉斯函数相结合,先检测多尺度的 Harris 特征点,然后用高斯拉普拉斯函数取局部极值剔除冗余点,缺点是计算量大.黄波等^[9]提出归一化直方图仿射不变量检测算法,从目标图像中提取直方图仿射不变量特征,对光照和视角变化具有良好的适应性,但是检测速度较慢.Leutenegger 等^[10]提出的 BRISK(Binary Robust Invariant Scalable Keypoints)算法,使用基于尺度空间的快速角点检测算法,在连续尺度空间内检验相邻区域灰度变化显著的点作为角点,再通过非极大值抑制以及尺度空间曲线插值来获得具有亚像素精度的特征点信息,速度和稳定性都较好.

对于特征点描述,为了使局部区域的特征能够对旋转、尺度缩放、光照和轻量级的仿射变换保持不变.研究者提出了很多局部特征点描述算法,常见的有以下一些:SIFT(Scale-Invariant Feature Transform)^[11]、SURF(Speeded Up Robust Features)^[12]特征描述对于图像变化都具有鲁棒性,但缺点在于描述符维数高,因而导致了后续搜索和匹配占用内存多而且速度较慢,无法直接移植到智能终端.最近提出的 BRIEF(Binary Robust Independent Elementary Features)^[13]、ORB(ORiented Brief)^[14]、FREAK(Fast Retina Keypoint)^[15]特征描述,都采用了二进制字符串作为描述符,降低了描述符的维度,无需考虑类似 SIFT 和 SURF 等高维特征矢量在欧氏空间的匹配复杂度.BRIEF 的缺点在于不具备旋转不变性、对噪声敏感、不具备尺度不变性.ORB 在 BRIEF 基础上增加了对特征主方向的计算,因此具有旋转不变性.FREAK 的描述符延续了 ORB 的思路,但在采样点的选择方法上不同,它使用了固定的采样点模式,采用预先定制的固定空间采样模式,有助于提高描述符的可重复性.

对于特征点匹配,一般采用某种代价函数或者距离函数来进行度量,它和特征空间密切相关.采用代价函数判断相似性的如 Ferns 算法^[16]通过对特征点进行后验概率判断,确定特征点所归属的类,需要大量的学习样本才能生成准确率比较高的判定树,并且在训练

的时候会占用大量内存,如果对于单个场景识别智能手机还能承受,但对于多个场景该类算法不仅运行速度较慢,还会影响其它程序.关于距离度量相似性的分为欧式距离度量和汉明距离度量.近似最近邻(Approximate Neighbor Nearest)方法如 K-means^[17],采用欧式距离度量,但是该方法在特征向量维度较高时,计算复杂度也在增大.Brisk^[10],Freak^[15]采用汉明距离度量,只需要简单的异或操作,速度非常之快,比较适合智能手机端的硬件条件.

根据以上所述,本文提出了一种适用于智能手机的识别算法,通过使用 Brisk 提点算法和 Freak 描述子解决了智能手机本身的处理能力和内存容量问题,再通过带重力方向的 Brisk 特征点降低了图像内部相似特征误匹配问题;最后,建立描述符多索引结构,完成相似描述符的快速查找.实验结果表明,该算法识别速度快,内存的占用量少,识别准确率也较高,适合移动平台的视觉搜索.

2 智能手机终端的特征点检测

2.1 智能手机平台

目前智能手机发展非常迅速,基本上都集成高速处理芯片、大容量存储器、内置了百万像素的摄像头和 WIFI 无线局域网接口卡,甚至一些高端的设备中还集成了大量传感器如 GPS 定位传感器和加速度传感器等.这些性能不断提高的硬件设备,给计算机视觉和图像处理领域带来了新的发展空间,许多原来运行在 PC 上的计算机视觉算法,通过改进和优化,能有效的运行的智能终端的平台上.

但是,相对于现在配备高速处理芯片和大容量存储芯片的服务器,智能手机还存在以下不足:较低的图像分辨率、缺少专门的图形加速器、没有浮点运算单元、尤其是低端处理器和低容量存储器等,限制了在智能手机上进行复杂运算和大规模计算.

针对智能手机存在的以上问题,本节后续部分的特征点检测和特征点描述算法分别从计算量和内存空间两方面进行考虑,使算法满足计算量小、占用空间低等特性,适合运行在智能手机上.

2.2 特征点检测

特征点就是图像局部区域中变化最剧烈的部分,本文采用提点速度快,计算量小的 BRISK 特征点检测算法,适合智能手机平台的同时兼顾了提点算法的实时性.

BRISK 检测算法由 Leutenegger^[10]提出,通过搜索图像尺度空间中非极大值抑制的角点来实现.BRISK 算法首先在图像各尺度层搜索突出性的角点,作为预备特征点;再对预备特征点集,在其尺度空间相邻的上下两

层的邻域内实行非极大值抑制,剔除一些非极大值的点,剩下的就是在尺度空间的邻域中具有独特性的唯一的特征点。

BRISK 算法使用 AGAST^[18]算法在离散尺度空间中的每一层搜寻角点,使用像素亮度作为比较指标,衡量待测点的突出性。突出性的度量为 FAST 得分,计算公式为:

$$S(x) = \begin{cases} 1, & I_c + \tau < I_x (x \in S_{\text{brightness}}) \\ 0, & I_c - \tau \leq I_x \leq I_c + \tau (x \in S_{\text{similar}}) \\ -1, & I_x < I_p - \tau (x \in S_{\text{darkness}}) \end{cases} \quad (1)$$

$$V = \max\left(\sum_{x \in S_{\text{brightness}}} |I_{p \rightarrow x} - I_p| - t, \sum_{x \in S_{\text{darkness}}} |I_p - I_{p \rightarrow x}| - t\right) \quad (2)$$

其中 V 为 FAST 响应得分, $S_{\text{brightness}}$ 为邻域点的亮度集合, S_{darkness} 为邻域点的暗度集合, I 为不同点的灰度, I_c 为中心像素点, I_x 为邻域像素点, τ 为衡量显著与否的灰度差阈值。如果中心点与邻域点的亮度差小于 τ , 便认为中心点与邻域亮度相近, 即不满足显著亮或暗的条件。

而邻域的选择则要根据应用平台以及需求来定, 常用的有 5-8、8-12、9-16 的模板。如图 1 所示, 在 AGAST_9-16 的模板下, 在待测点同层的 16 个点组成的邻域中, 当 FAST 响应得分不小于 9 (即至少有 9 个邻域点的灰度全大于或者全小于待测的中心点的亮度) 时, 便认为中心点在邻域中具有突出性, 将纳入预备特征点中。

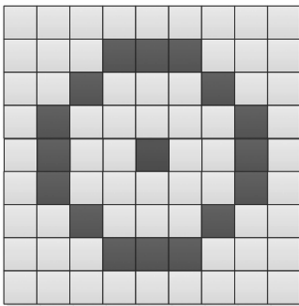


图1 AGAST_9-16模板

将尺度空间中所有搜索到的角点作为预备特征点, 再对所有预备特征点在尺度邻层实施非极大值抑制。最后被确认为特征点的像素点 P_σ , 它在尺度空间的上下邻层中对应的点 $P_{\sigma-1}$ 和 $P_{\sigma+1}$ 在其各自的尺度层上也应该具有突出性, 而且 P_σ 的 FAST 响应得分应该比 $P_{\sigma-1}$ 和 $P_{\sigma+1}$ 都要高, 确保特征点在尺度空间中的独特性。

2.3 Gravity-aware BRISK 特征点

由于直接使用 BRISK 对拍摄图像进行匹配效果并

不理想, 尤其是拍摄建筑物等存在大量相似结构的物体, BRISK 算法无法区分图像里面的相似结构, 本文中没有使用原始 BRISK 特征信息。文献[19]描述在场景内部存在相似特征或接近一致的特征, 单靠从图像像素强度得出的梯度主方向方向, 将难以区分相似特征点的误匹配问题。如图 2 左边部分, 场景的四个角区域, 对应于 4 个不同的自然特征, 但是按照特征点像素强度主方向进行规范化后, 变成了相同的特征。这种现象明显增加了特征点的误匹配率, 也降低了描述符的区分能力。智能手机上的众多的传感器为提取更有区分能力的图像特征点提供了条件, 利用重力在一定程度能够解决上述的问题。如图 2 右边部分, 按照重力方向对齐后, 4 个局部特征的归一化区域之间的差异更加明显。

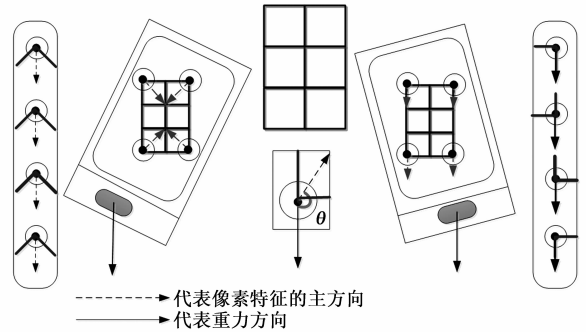


图2 特征点像素强度主方向与重力方向的夹角

所以, 将重力方向加入特征点的描述中, 明显的解决了自然特征的匹配问题。本文提出 Gravity-Aware BRISK 的特征点, 将摄像机坐标系中的重力方向依据摄像机成像模型投射到图像平面, 得到特征点的重力方向, 对特征点沿重力方向对齐, 计算重力方向与特征点邻域像素强度的主方向之间夹角, 将其作为特征点的一个描述。原理如图 2 下面部分所示。带重力方向的特征点匹配时, 首先进行夹角相似性判定, 设定一个阈值, 判断夹角差的绝对值小于这个阈值; 再进行传统的最短距离查找。它在一定程度上克服了传统 BRISK 算法对于相似结构的特征点难以区分的缺点, 也节省了描述符匹配时间, 在性能上优于传统的 BRISK 算法。

普通的智能手机带有: 加速度传感器、姿态传感器和陀螺仪 (gyroscope) 等传感设备。加速度传感器为每一帧图像提供了重力方向, 用三维的矢量表示, 如 $\mathbf{g} = [g_x, g_y, g_z]^T$, 通常在摄像机坐标系中对 \mathbf{g} 进行归一化 $\|\mathbf{g}\| = 1$, 并将重力矢量投影到图像所在二维平面上, 通过式(3)计算投影平面的图像像素点 p 重力方向图像中点 $\mathbf{p} = [u, v, 1]^T$ 处的重力方向向量为

$$\mathbf{d} = \mathbf{p}' - \mathbf{p} \quad (3)$$

其中 $\mathbf{d} = [d_u, d_v, 0]^T$ 为点 \mathbf{p} 处的重力方向, \mathbf{p}' 由下式计算得出:

$$\mathbf{p}' = [wu', ww', w]^T = \mathbf{p} + \mathbf{K}\mathbf{g} \quad (4)$$

其中 $\mathbf{K} = \begin{bmatrix} f_u & 0 & p_u \\ 0 & f_v & p_v \\ 0 & 0 & 1 \end{bmatrix}$ 为手机相机的内参矩阵. \mathbf{d} 也可

以通过一种计算量较小的近似方法求得:

$$\mathbf{d} \propto [g_x(p_u - u) + f_u g_x, g_x(p_v - v) + f_v g_x, 0]^T$$

由 \mathbf{d} 可计算出图像中像素点重力方向的角度为

$$\theta_g = \arctan\left(\frac{d_v}{d_u}\right) \quad (5)$$

其中 $[p_u, p_v]^T$ 是主点坐标, f_u, f_v 是水平与垂直方向的焦距.

3 特征点的描述

本文选用 FREAK 描述符, 而没有选择 SIFT、SURF、BRISK 等描述符, 在文献[15]中阐述了前者计算速度相比后三种更快、内存负载更低的, 鲁棒性更好. FREAK^[15] 是一种二值描述符, 采用固定的采样模式, 以特征点为中心的较近邻域内放置的采样点密度较大, 而且每个采样点的高斯核参数较小; 而离特征点较远的边缘区域, 采样点的密度小而且高斯核参数较大. 这与我们平日感受中的距离视野中心越远的物像越模糊的主观经验相符.

FREAK 特征主方向的计算方法与 BRISK 类似, 都使用局部梯度来表示特征主方向, 但不同之处在于计算梯度的采样点对的选取, 后者使用长距离的采样点对, 而前者则选用如图 3 所示, 简单的相对于中心点对称的几组采样点对.

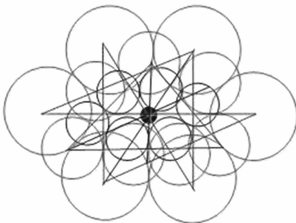


图3 确定FREAK特征方向的采样点对

特征主方向的梯度 $\mathbf{grad} = \begin{pmatrix} grad_x \\ grad_y \end{pmatrix}$ 的计算公式如下:

$$\mathbf{grad} = \frac{1}{M} \sum_{(P_j, P_i) \in C} (I(P_j, \sigma_j) - I(P_i, \sigma_i)) \frac{P_j - P_i}{\|P_j - P_i\|^2} \quad (6)$$

其中 $I(P_j, \sigma_j)$ 和 $I(P_i, \sigma_i)$ 分别代表高斯平滑后的一对采样点的亮度, C 是用来计算梯度的采样点对总集, M 为 C 中采样点对数, P_j 和 P_i 为一对采样点的空间坐标矢量.

求出主方向后, 把采样模式点集沿特征点 \mathbf{K} 旋转

$\theta_g = \arctan^2(g_y, g_x)$, 开始构建描述符 D :

$$D = \sum_{0 \leq e \leq N} 2^e T(P_i^e, P_j^e) \quad (7)$$

(P_i^e, P_j^e) 为特征点 \mathbf{K} 沿主方向对齐后, 邻域内的一对采样点, N 是描述子长度. 而 T 满足:

$$T(P_i^e, P_j^e) = \begin{cases} 1, & I(P_i^e) - I(P_j^e) > 0 \\ 0, & \text{其他情况} \end{cases} \quad (8)$$

FREAK 使用 43 个采样点, 生成 $43 \times 42 = 1118$ 比特的描述符, 通过比较点总方差最大的标准对采用点对进行筛选, 最终的确定 FREAK 描述符只有 512 比特的长度.

4 特征点匹配

传统的描述符匹配算法都是将描述符与所有的样本描述符进行比较, 来确定相似描述符. 在样本图像较多的情况下, 传统算法的描述符匹配时间成线性增长, 对实时性要求较高的图像识别系统是不能接受的. 本文实现了一种相似描述符的快速查找方法, 只用比较部分描述符.

4.1 相似描述符的快速查找

首先, 对 512 位的 FREAK 描述符用图 4 所示的方式切分为多个互不相交子段描述符, 每个子段长为 8 位, 一共 64 个子段; 再对子段描述符建立如图 5 所示的二级索引. 图 5 的一级索引为子段号, 二级索引为 8 位的二进制向量 (00000000 ~ 11111111) 对每个样本描述符先进行分段, 每个子段根据子段号和子段描述符对应的二进制索引找到相应的倒排链表, 将描述符标识和所在图像标识生成一个节点插入到链表中. 但是, 为了节省存储空间和方便后续进行与操作, 本文采用图 6 中的方式将所有样本图像的描述符存储在一个数组中, 图像对应描述符在数组中的开始和结束序号与该图像进行绑定, 将数组中元素序号作为描述符 ID 号插入到倒排链表中. 这样通过序号不仅可以快速找到描述符, 而且可以定位对应的图像.

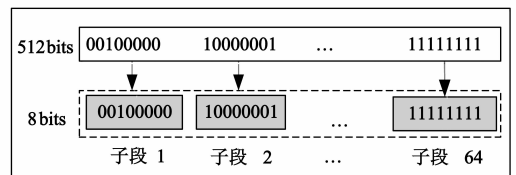


图4 512位二进制描述符的分段

根据以上思想, 优化原有二级索引的倒排链表, 以索引值为聚类中心, 满足汉明距离值小于等于 $\left\lfloor \frac{R_{th}}{n} \right\rfloor$ 都插入到倒排链表中. 实现相似描述符查找时, 先将查询描述符进行分段, 找到每个子段对应的倒排链表, 再

对链表中的 ID 集合进行与操作,如式(9). 设子段 1,子段 2,⋯,子段 n , 对应倒排链表的集合为 D_1, D_2, \dots, D_n, U 为子段描述符对应 ID 集合的交集.

$$U = D_1 \cup D_2 \cup \dots \cup D_n \quad (9)$$

最后,将查询描述符与 U 中 ID 对应的描述符进行逐个比较,找出满足式(10)样本描述符为相似描述符.

$$(|\alpha_i - \beta_j| < \theta_{th}) \&\& (r_{min} < R_{th}) \quad (10)$$

α_i, β_j 分别为查询描述和样本描述符重力方向与像素强度主方向的夹角, θ_{th} 为夹角阈值, r_{min} 为 U 中样本描述符与查询描述符最近的距离, R_{th} 为汉明阈值.

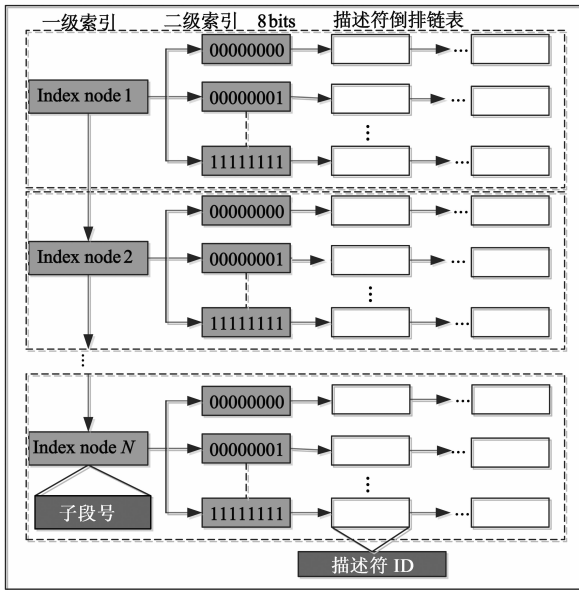


图5 样本描述符的二级索引表

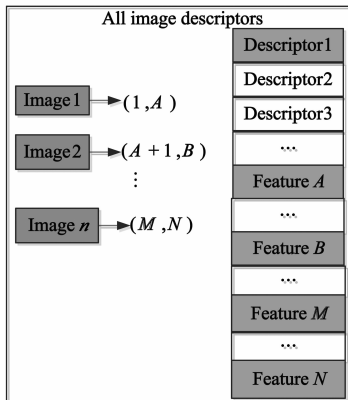


图6 样本描述符的存储方式

4.2 识别目标确定

当完成查询图像的描述符匹配以后,需要判断相似特征点的数目占样本图像和查询图像本身的比重是否大于设定的阈值来决定是否查找到成功匹配的样本.

计算查询图像 A 的特征向量集与训练图像库中所

有特征向量集的匹配点数目,选定匹配点数目最多的一张图像 B ,测试匹配率,当匹配率大于设定一个阈值为匹配成功的图像.

设匹配率阈值为 M_{th} , A 与 B 的特征点数目为 k_1 和 k_2 , 匹配点数目为 m . 如果满足下式:

$$\frac{m}{k_1} > M_{th} \&\& \frac{m}{k_2} > M_{th} \quad (11)$$

则说明两张图像的匹配程度较高,匹配成功. 该样本图像即为对应的匹配目标.

R_{th} 和 M_{th} 的不同取值会造成识别效果的差异,将在实验部分给出实验数据以决定最好的取值组合.

5 实验验证

5.1 实验条件

实验用到的智能手机为: HTC (919d), CPU 频率 (1.9GHz), RAM(2GB), 操作系统为 Android 4.2, 外用存储设备为 32G 的 MicroSD.

实验数据为两组数据集: 一组为我们用智能手机采集的带有重力方向信息的 1000 个室外场景图像, 每个场景采集 5 个不同视角的图像, 一共 5000 副图像, 部分示例图像, 如图 7 所示; 另一组国际图像标准库 UK-Bench^[19], 不带有任何传感器信息, 仅包含 2550 个不同的物体, 每个物体包括 4 个从不同的角度拍摄的图像, 一共 10200 幅图像, 如图 8 所示.



图7 智能手机采集的数据图像

5.2 实验过程

算法的流程如图 9 所示, 分为训练阶段和在线识别阶段. 训练阶段将所有的图像进行降分辨率处理, 统一转换为 320×240 . 对智能手机采集带有重力方向的户外图像集, 提取 brisk 特征点, 生成 FREAK 特征描述符; 选取前 100 个响应值较大特征点计算像素强度主方向与重力方向的夹角; 最后, 建立多级索引表, 并将索引表信息存储到文件中. 下次使用该算法时, 只需要将索引表载入内存, 就可以进行识别, 减少对样本的重复性训

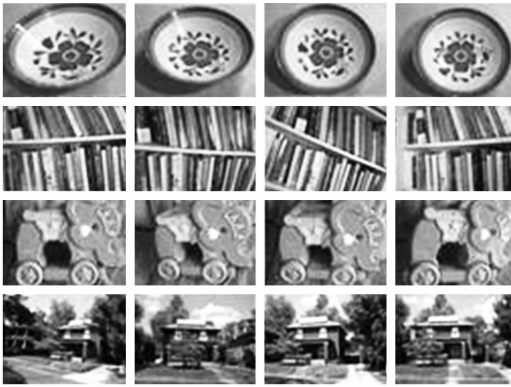


图8 UKBench 数据集

练. 而对不带有重力信息的 UKBench 库中图像, 仅计算特征描述符, 将夹角全都设置为 0. 在线识别阶段: 首先通过智能手机采集图像的视觉信息和重力方向信息; 再提取特征点, 计算特征描述符和特征方向与重力主方向的夹角, 然后进行图像匹配.

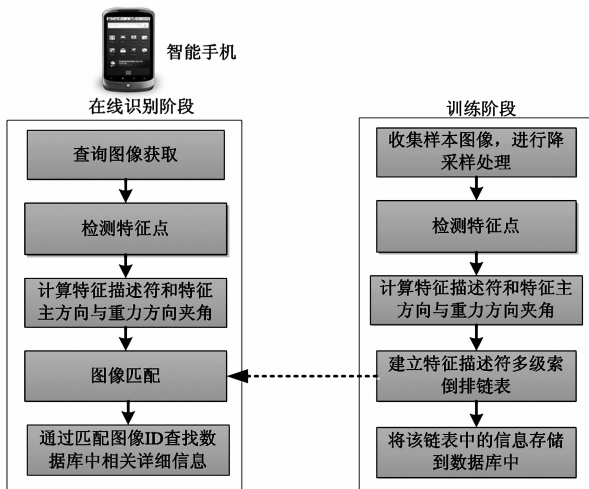


图9 本文算法的流程图

5.3 算法性能

在图像识别算法中, 通常从训练时间、训练内存占用、识别时间和识别精度四个指标来衡量算法的性能.

(1) 训练时间是从特征点提取到描述符建立消耗的时间.

(2) 训练内存占用是训练阶段内存消耗的最大值.

(3) 识别时间是从开捕获视频开始到智能终端出现结果的时间.

(4) 识别精度为智能终端成功识别的次数与总识别次数的比值.

本文在智能手机上实现了提出的算法, 并从上面提出的四个方面对算法进行对比. 每次实验我们随机选取不同的场景, 每个场景的所有图像为训练样本. 实验结果取 10 次测试的平均值. 测试方法为开启智能手

机摄像头读入查询图像, 提取特征信息, 在样本库中查找相似图像; 如果返回的相似图像与待识别图像属于同一场景的便视为识别成功, 否则视为失败.

5.3.1 特征点检测时间和数量对比

图 10 为任意拍摄一张室外场景图像进行降分辨率处理, 生成三张分辨率为 640×480 , 320×240 , 160×120 的图像. 表 1 是本文算法对三张图像的进行了特征点计算的时间和检测特征点的数量, 可以看出提点和描述符的总时间是非常短的.

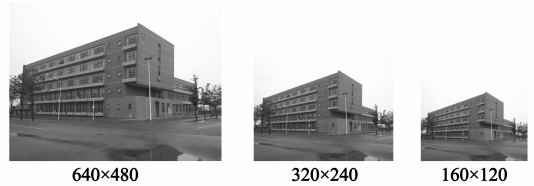


图10 三种不同分辨率的测试图像

表 1 不同分辨率图像的特征点检测

分辨率	提点时间(ms)	描述时间(ms)	特征点数
640×480	21.05	26.83	916
320×240	17.63	21.08	653
160×120	13.06	18.72	370

5.3.2 图像训练占用时间和内存对比

图 11 是算法训练所占用时间和内存对比. 从图 11 (a) 得出, 本文的算法在最大样本图像数量时占用内存最大值不超过 30M, 相对于 2G 内存的智能手机, 不影响其他程序性能. 从图 11 (b) 得出, 本文算法的样本图像的训练时间为 200s 左右, 能迅速完成图像的训练. 实验结果证明, 本文算法训练图像的内存占用和时间消耗都较小适用于移动终端.

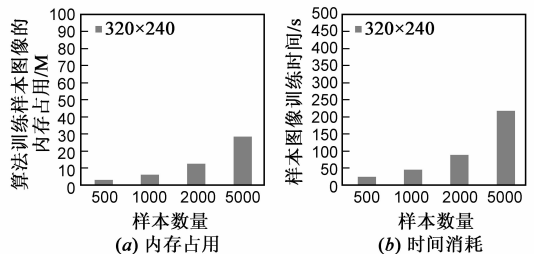


图11 训练样本图像占用内存和时间

5.3.3 图像识别时间和精度对比

表 2, 图 12 是在智能手机采集图像数据集上的测试结果. 表 2 是在 300 个场景, 1500 张样本图像下, 进行图像识别实验统计出的算法各个步骤的平均消耗时间, 可以看出查询一张图像的时间非常短, 小于 300ms, 表明本文算法实时性较好, 因为依靠智能手机本身的识别算法相比传统的 C/S 架构的视觉算法来说, 减少

了通过无线网络传输数据的延时.图 12 是本文算法取不同参数阈值在不同规模图像库上识别精度的比较.在样本图像较少时,算法都有较好的识别精度;在样本图像增大到 1000 时,算法的识别率都保持在 70% 以上;但是随着样本库的持续增大,识别精度都有下降的趋势,同时汉明距离阈值的增加,误匹配点也在增加,使识别精度有所下降.匹配率阈值对识别精度也有影响,在室外建筑变化较大的图像特征匹配点数目较少的情况下,设置较高的匹配率阈值反而是很多正确匹配的图像被过滤掉.本文实验结果表明,本文算法在汉明阈值 R 为 40,匹配的特征夹角差的阈值 θ 为 30,比率 M 为 20% 效果最好.

表 3 UKBench 数据集图像测试结果

UKBench	100		500		1000		2000	
	时间(ms)	精度(%)	时间(ms)	精度(%)	时间(ms)	精度(%)	时间(ms)	精度(%)
本文算法	65.28	82.6	168.42	76.3	274.06	63.5	632.91	56.8

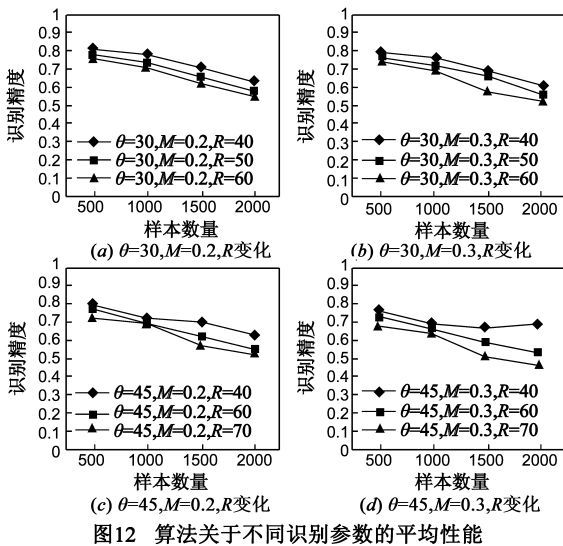


图 12 算法关于不同识别参数的平均性能

6 结论与展望

本文提出了一种适用于智能手机上图像识别算法,解决了在智能手机上实现大规模图像识别的问题.在特征提取方面,提出带重力方向的 BRISK 特征点和用 FREAK 进行描述;在目标匹配方面,首先使用方向夹角过滤相似特征的误匹配点,提高了识别精度,再通过判断匹配特征点占查询图像和样本图像的双比重,来确定是否成功匹配,进一步保证了识别的准确性.实验结果证明,该识别算法可在智能手机平台上有效运行.本文算法也有不足之处,没有对描述符进行压缩和利用特征点的空间关系优化匹配算法.下一步的研究工作是实现描述符的高比率压缩和利用特征点空间关系重排匹配结果.

表 2 查询图像的平均搜索时间

执行步骤	时间(ms)
特征点检测	21.08
特征点描述	2.17
匹配	218.36
总计	241.61

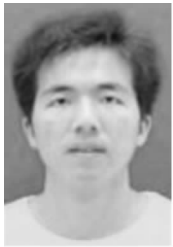
表 3 是在标准图像识别库 UKBench 数据集上的测试结果.从表 3 得出本文算法的识别时间始终保持在 1s 以内,不影响用户的体验.因此,通过在公开的标准数据集上的实验再次验证了本文算法的有效性,既能进行快速识别,同时保持较高的识别精度.

参考文献

- [1] D Chen, G Baatz, K Koeser. City-scale landmark identification on mobile devices[A]. Proceedings of 24th IEEE International Conference on Computer Vision and Pattern Recognition[C]. Piscataway: IEEE, 2011. 737 - 744.
- [2] Xiaoyu Wang, MingYang, Timothee Cour. Contextual weighting for vocabular Tree based image retrieval[A]. Proceedings of the 13th IEEE International Conference on Computer Vision [C]. Piscataway: IEEE, 2011. 209 - 216.
- [3] 赵永威, 郭志刚, 李弼程, 等. 基于随机化视觉词典组和上下文语义信息的目标检索方法[J]. 电子学报, 2012, 40(12): 2473 - 2479.
Zhao Yong-wei, Guo Zhi-gang, Li Bi-cheng, et al. Object retrieval method based on randomized visual dictionaries and contextual semantic information[J]. Acta Electronica Sinica, 2012, 40(12): 2473 - 2479. (in Chinese).
- [4] R Muralidharan, C Chandrasekar. 3D object recognition using multiclass support vector machine-k-nearest neighbor supported by local and global feature[J]. Journal of Computer Science, 2012, 8(8): 1380 - 1388.
- [5] 袁杰, 魏宝刚, 王李冬. 一种综合 PHOG 形状和小波金字塔能量分布特征的图像检索方法[J]. 电子学报, 2011, 39(9): 2115 - 2119.
Yuan Jie, Wei Bao-gang, Wang Li-dong. An image retrieval method synthesizing PHOG shape description and wavelet pyramid energy distribution[J]. Acta Electronica Sinica, 2011, 39(9): 2115 - 2119. (in Chinese).
- [6] Wagner D, Reitmayr G, Mulloni A, et al. Real-time detection and tracking for augmented reality on mobile phones[J]. IEEE Transaction on Visualization and Computer Graphics, 2010, 16

- (3):355 – 368.
- [7] Jae-Pil Heo, YoungWoon Lee, Junfeng He, et al. Spherical hashing[A]. Proceedings of the 25th IEEE International Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2012. 2957 – 2964.
- [8] Mikolajczyk K, Schmid C. Scale and affine invariant interest point detectors[J]. International Journal of Computer Vision, 2004, 60(1): 63 – 86.
- [9] 黄波, 赵继印, 郑蕊蕊, 等. 基于多尺度自卷积归一化直方图的仿射不变量模式识别[J]. 电子学报, 2011, 39(1): 65 – 68.
Huang Bo, Zhao Ji-yin, Zheng Rui-rui, et al. Affine invariant pattern recognition based on multi-scale autoconvolution normalized histograms[J]. Acta Electronica Sinica, 2011, 39(1): 65 – 68. (in Chinese).
- [10] Leutenegger S, Chli M, Siegwart R Y. BRISK: Binary robust invariant scalable keypoints[A]. Proceedings of the 13th IEEE International Conference on Computer Vision [C]. Piscataway: IEEE, 2011. 2548 – 2555.
- [11] David G Lowe. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [12] Herbert B, Tinne T, et al. SURF: speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346 – 359.
- [13] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features[A]. Proceedings of the 11th European Conference on Computer Vision [C]. Heidelberg: Springer, 2010. 778 – 792.
- [14] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to sift or surf[A]. Proceedings of the 13th IEEE International Conference on Computer Vision [C]. Piscataway: IEEE, 2011. 2564 – 2571.
- [15] Alahi A, Ortiz R, Vanderghyest P. Freak: Fast retina keypoint [A]. Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2012. 510 – 517.
- [16] Ozuysal M, Calonder M, Lepetit V. Fast keypoint recognition using random ferns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 448 – 461.
- [17] Ramasubramanian V, Paliwal K. Fast k-dimensional tree algorithms for near-neighbor search with application to vector quantization encoding [J]. IEEE Transaction on Signal Processing, 1992, 4(3): 518 – 531.
- [18] E Mair, G D Hager, D Burschka, M Suppa, et al. Adaptive and generic corner detection based on the accelerated segment test [A]. Proceedings of the 11th European Conference on Computer Vision [C]. Heidelberg: Springer, 2010. 183 – 196.
- [19] D Kurz, S. Benhimane. Inertial sensor-aligned visual feature descriptors [A]. Proceedings of 24th IEEE Conference on Computer Vision and Pattern Recognition [C]. Piscataway: IEEE, 2011. 161 – 166.
- [20] Norouzi M, Punjani A, Fleet D J. Fast search in Hamming space with multi-index hashing[A]. Proceedings of the 25th IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE, 2012. 3108 – 3115.

作者简介



桂振文 男, 1983 年出生于湖南祁阳县. 现在北京理工大学计算机学院攻读博士学位. 主要从事计算机视觉、模式识别和图像处理方面的研究.

E-mail: quizhenwen1983@bit.edu.cn



刘越 男, 1968 年出生于吉林省长春市. 2000 年获得吉林大学通信与信息系统专业的博士学位. 现任北京理工大学光电学院教授、博士生导师. 主要从事虚拟现实与增强现实方面的教学和研究工作.

E-mail: liuyue@bit.edu.cn