



(12) 发明专利

(10) 授权公告号 CN 108959478 B

(45) 授权公告日 2021.06.22

(21) 申请号 201810642032.X

G06K 9/46 (2006.01)

(22) 申请日 2018.06.21

G06K 9/62 (2006.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 108959478 A

(56) 对比文件

(43) 申请公布日 2018.12.07

邹琴. 移动云环境下密文图像检索技术研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2018,7-50.

(73) 专利权人 中南林业科技大学
地址 410000 湖南省长沙市雨花区韶山路498号

王艳. 基于点特征的立体匹配算法研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2009,31.

(72) 发明人 秦姣华 李浩 向旭宇 潘丽丽
谭云 马文涛

徐望明. 面向图像检索和分类任务的稀疏特征学习.《中国博士学位论文全文数据库信息科技辑》.2014,31-37.

(74) 专利代理机构 深圳市深软翰琪知识产权代理有限公司 44380

邹琴. 移动云环境下密文图像检索技术研究.《中国优秀硕士学位论文全文数据库信息科技辑》.2018,7-50.

代理人 黄美成

审查员 熊菡

(51) Int. Cl.

G06F 16/583 (2019.01)

权利要求书3页 说明书16页 附图5页

(54) 发明名称

一种云环境下的密文图像检索方法及系统

(57) 摘要

本发明公开了一种云环境下的密文图像检索方法及系统,首先,从自适应阈值和特征点预选筛选两方面优化Harris算法,并提取图像特征。其次,采用SURF算法和词袋模型生成每一幅图像的Harris角点特征向量。然后,采用局部敏感哈希(Locality Sensitive Hashing, LSH)算法对特征向量构建可搜索索引,并用传统加密方案对图像以及索引进行加密,最后,在云服务器上进行安全的相似性检索。实验结果证明,通过对Harris角点优选及SURF与词袋模型的特征描述,并对局部敏感哈希算法的参数进行了优化,本发明提出的检索方案与现有加密检索方案对比,不仅缩短了特征提取时间,而且有效提高了加密图像检索效率。



1. 一种云环境下的密文图像检索方法,其特征在于:

(一) 针对数据拥有者:

对于一个图像集M,实施以下操作:

(1) 用 $Gen_{Harris}(Harris, M)$ 生成特征集为 $\{G_i\}_{i=1}^n$;

(2) 用 $Gen_{feature}(\{G_i\}_{i=1}^n)$ 生成特征向量 $\{f_i\}_{i=1}^n$;

(3) 用 $Build_{index}(\{f_i\}_{i=1}^n)$ 生成索引I,用 $Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 分别生成加密特征向量 $\{f'_i\}_{i=1}^n$ 、加密图像集M'、加密索引I';

(4) 将 $\{f'_i\}_{i=1}^n$ 、M'、I'发送到云服务器,将加密密钥K发送给查询用户;

$Gen_{Harris}(Harris, M)$ 表示采用Harris算法对图像M提取图像特征;

$Gen_{feature}(\{G_i\}_{i=1}^n)$ 表示对特征集 $\{G_i\}_{i=1}^n$ 生成特征向量;

$Build_{index}(\{f_i\}_{i=1}^n)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 构建索引;

$Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 、图像集M、索引I进行加密;

(二) 针对查询用户:

对于查询图像集 M_q ,执行以下操作:

(1) 用 $Gen_{Harris}(Harris, M_q)$ 生成特征集为 $\{G'_i\}_{i=1}^{n_q}$;

$Gen_{Harris}(Harris, M_q)$ 表示采用Harris算法对图像 M_q 提取图像特征;

(2) 用 $Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$ 生成特征向量 $\{f_{qi}\}_{i=1}^{n_q}$;

$Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$ 表示对特征集 $\{G'_i\}_{i=1}^{n_q}$ 生成特征向量;

(3) 用 $Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$ 生成陷门TD,然后将陷门TD发送到云服务器;

$Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$ 为产生陷门的函数;

(4) 用解密函数 $Dec_{data}(K, \mathcal{R})$ 解密返回相似图像 \mathcal{R} ;

(三) 针对云服务器:

用 $Search(I, M', \{f'_i\}_{i=1}^n, TD)$ 算法进行检索并返回相似结果集 \mathcal{R} ;

$Gen_{Harris}(\cdot)$ 是指基于自适应阈值与Forstner方法进行Harris角点优选;

确定候选节点后,再根据最大角点响应函数对候选点进行筛选,确定提取的预筛选特征点总数 c_1 ,最后结合Frostner算法确定最佳候选点总数 c_2 ,实现Harris角点优选;

Harris角点优选包括以下步骤:

步骤1:采用8邻域相似像素分析法确定候选集C;

对于图像中的任一目标像素点 (x, y) ,计算该目标像素点与8邻域范围内像素点灰度差的绝对值 Δ ,通过与设定的阈值 t 相比较来确定是否相似,统计目标像素点与周围8个点的相似个数 $N(x, y)$,如下式所示:

$$N(x, y) = \sum_{i, j} \chi(x + i, y + j) (-1 \leq i \leq 1, -1 \leq j \leq 1, \text{且 } i \neq 0, j \neq 0)$$

式中:识别函数 $\chi(x+i, y+j) = \begin{cases} 1, & \Delta(x+i, y+j) \leq t \\ 0, & otherwise \end{cases}$, 当 $2 \leq N(x, y) \leq 6$ 时, 将该目标

像素点 (x, y) 视为候选点, 用 C 表示候选点集合;

步骤2: 计算每个候选点的响应函数CRF, 定义阈值 \mathcal{T} 为最大CRF值的 ρ 倍, 即

$$\mathcal{T} = \rho * CRF_{max}$$

根据最大角点响应函数 \mathcal{T} 对候选点进行筛选, 确定提取的预筛选特征点总数 c_1 , 预筛选特征集 C_1 ;

步骤3: 结合Frostner算法确定最佳候选点集 $G (G = \{(x_i, y_i)\}_{i=1}^{c_2})$ 以及最佳候选点总数 c_2 ;

首先, 以预筛选特征集 C_1 任意一点 (x_1, y_1) 为中心建立 $3*3$ 窗口, 对该窗口内的每个点计算协方差矩阵cov:

$$cov = \begin{bmatrix} \sum J'_x{}^2 & \sum J'_x J'_y \\ \sum J'_x J'_y & \sum J'_y{}^2 \end{bmatrix};$$

其中, J'_x, J'_y 是Robert梯度算子;

$$J'_x = f(x+1, y+1) - f(x, y), \quad J'_y = f(x+1, y) - f(x, y+1);$$

f 是灰度函数, $f(x, y)$ 表示点 (x, y) 的灰度值;

接着, 计算特征点的权值 ω 和圆度 τ ;

$$\omega = \frac{\det(cov)}{\text{trace}(cov)}$$

$$\tau = \frac{4\det(cov)}{(\text{trace}(cov))^2}$$

其中 $\det(cov)$ 是协方差矩阵cov的行列式, $\text{trace}(cov)$ 是协方差矩阵cov的迹; 然后, 将 ω, τ 分别与给定阈值 $\mathcal{T}_\omega, \mathcal{T}_\tau$ 比较, 将满足 $\omega > \mathcal{T}_\omega$ 且 $\tau > \mathcal{T}_\tau$ 的备选点加入特征集 C_2 ; 最后在窗口内, 依据权值 ω 将满足条件 $\omega(x, y) = \max\{\omega(x, y)\}$ 的点加入最佳候选点集 G , 最佳候选点集 G 中的候选点的个数即为 c_2 。

2. 根据权利要求1所述的云环境下的密文图像检索方法, 其特征在于, $\text{Gen}_{\text{feature}}(\cdot)$ 是指结合SURF算法描述Harris特征点, 并结合词袋模型生成图像的特征向量。

3. 根据权利要求1所述的云环境下的密文图像检索方法, 其特征在于, $\text{Build}_{\text{index}}(\cdot)$ 是指构建哈希索引; 采用基于 ρ 稳定的LSH函数族构建哈希表, 作为哈希索引。

4. 根据权利要求2所述的云环境下的密文图像检索方法, 其特征在于: 生成图像的特征向量的步骤如下:

步骤s1: 用 \mathcal{K} -mean聚类算法对局部特征 $\{C_i\}_{i=1}^{c_2}$ 进行聚类, 形成一个视觉单词; $\{C_i\}_{i=1}^{c_2}$ 是 $G = \{(x_i, y_i)\}_{i=1}^{c_2}$ 在聚类中的表示;

步骤S1a, 随机选择 \mathcal{K} 个点作为聚类中心 $\{C_1, C_2, \dots, C_{\mathcal{K}}\}$; 然后, 用下式计算特征集 G 中的每个数据点到这 \mathcal{K} 个聚类中心的距离 d , 并将数据点按距离分配到最近的聚类中心, 形成 \mathcal{K}

个簇 $\mathbf{u} = \{\mathbf{u}_i\}_{i=1}^{\mathcal{K}}$;

$$\delta = \sum_i^{c_2} \sum_j^{\mathcal{K}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

其中, (x_i, y_i) 是 C_2 中的点, (x_j, y_j) 是 \mathcal{K} 个点作为聚类中心中的点;

步骤S1b, 用下式计算簇的平均值 \mathbf{u} , 指定这些值为新的聚类中心, 即视觉单词 $\mathbf{u} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\mathcal{K}}\}$;

$$\bar{\mathbf{u}}_j = \frac{1}{|\mathbf{u}|} \sum_{j=1}^{\mathcal{K}} \mathbf{u}_j$$

其中, $\bar{\mathbf{u}}_j$ 表示第 j 个视觉单词 \mathbf{u}_j 的平均值, 即 \mathbf{u}_j 的总和与簇 \mathbf{u} 的个数的比值, \mathbf{u}_j 的总和是指簇中的特征点的特征值相加, $|\mathbf{u}|$ 是 \mathbf{u} 的个数;

步骤S1c: 重复以上步骤S1a和S1b, 直到聚类中心的值满足最小化均方误差MSE函数收敛; 此时视觉单词表示为 $\mathbf{u}' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_{\mathcal{K}}\}$; 每个 \mathbf{u}' 是一个 \mathcal{K} 特征向量;

$$MSE = \sum_{j=1}^{\mathcal{K}} \sum_{\mathbf{u}_i \in \mathbf{u}} |\mathbf{u}_i - \bar{\mathbf{u}}_j|^2;$$

步骤S2: 得到视觉单词之后, 将局部特征 C_i 按下式对应到视觉单词中;

$C_i = \omega' = \{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\}$; 其中 $\{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\}$ 表示 C_i 对应到视觉单词 \mathbf{u}' 的权重;

统计整幅图像的视觉单词出现频率, 视觉单词出现频率即统计每个视觉单词在局部特征出现的个数, 生成图像特征向量 f_i , 所有图像的特征向量表示为 $\{f_i\}_{i=1}^n$ 。

5. 根据权利要求3所述的云环境下的密文图像检索方法, 其特征不在于, 采用基于 ρ 稳定的LSH函数族构建哈希表作为哈希索引的步骤如下:

数据拥有者选取 L 个LSH函数 $\{h_1, h_2, \dots, h_L\}$, 并对所有的特征向量 $\{f_i\}_{i=1}^n$ 应用函数 $\mathcal{G}(f_i) = (h_1(f_i), \dots, h_L(f_i))$; $\mathcal{G}(f_i)$ 表示哈希函数族; 为了提高准确率, 重复这个过程 λ 遍, 从而生成 λ 个哈希表; 用 $\{D_{i,j}\}$, $i \in [1, \lambda]$, $j \in [1, N_i]$ 表示由局部敏感哈希函数生成的桶值集, N_i 表示第 i 个哈希表中的总桶数; $ID(m_t)$ 表示图像 m_t 将自身的ID关联至相应的桶值 $D_{i,j}$, 形成加密哈希表。

6. 一种云环境下的密文图像检索系统, 其特征不在于, 包括数据上传终端、云服务器和查询终端;

(1) 数据上传终端用于对图像拥有者将图像进行处理, 将处理后的图像的加密特征上传到云服务器;

(2) 云服务器用于存储数据上传终端上传的图像以及图像的加密特征; 云服务器还用于执行检索操作, 将检索结果返回到查询终端;

(3) 查询终端用于查询用户输入待查询的图像, 以及用于显示云服务器返回的查询结果;

采用权利要求1-5任一项所述的云环境下的密文图像检索方法实施图像检索。

一种云环境下的密文图像检索方法及系统

技术领域

[0001] 本发明涉及一种云环境下的密文图像检索方法及系统。

背景技术

[0002] 图文检索定义：

[0003] 将需要保护的隐私图像加密后上传到云服务器，这里上传的是加密后的图像集，检索的时候是根据用户的需求，用户可能输入一幅图像，然后在云端进行相应的检索，找到符合要求的图像反馈给用户，反馈的图像可以是多幅，按相似度进行排序，可以设定反馈回用户的图像的个数（也就是最相似的图像数量）。

[0004] 密文图像检索意思是对明文图像也就是普通图像进行加密，特别是对隐私图像，对隐私图像加密后没有授权的用户即使获得这个图像，打开图像也是乱码。

[0005] 云计算环境下密文图像检索是解决大规模图像存储管理困难与图像安全问题的关键，已得到各国政府、企业、研究人员的广泛关注和高度重视。如何使图像拥有者在云计算环境中加密存储自己的图像，又能让授权用户快速检索到自己需要的图像，是一个极具挑战性的课题。但这方面的研究仍处于起步阶段，有些关键问题还没得到有效解决。

[0006] 研究意义：

[0007] 大数据时代的到来使得图像数量呈几何指数增长，图像在医学、教育等领域发挥越来越重要作用。随着这些领域图像数量的增多，对于图像的高效存储和检索服务的需求得到加强。云计算的兴起，为用户提供按需购买、按用付费的计算和存储服务，成为图像存储和图像搜索服务外包的首要选择。通过将图像的搜索服务外包，数据拥有者不用在本地维护海量的图像数据库。查询用户不用跟图像拥有者进行交互，就可以直接通过云服务商搜索相似图像。云环境下多用户方便快捷的图像检索，带来了图像数据的安全性问题。为提高密文图像检索的有效性和速度，密文图像检索成为新的研究课题。

[0008] 随着成像传感器和手持电子设备的飞速发展，图像等多媒体数据在医药、杂志、广告、教育、娱乐等行业中扮演越来越重要的角色，且呈现爆炸性增长的趋势。截止2017年2月，人们在Flickr上传了超过130亿张图片；而Facebook的图片总量在2013年9月就已经达到2500亿，并且还在以每天3.5亿，即每秒4000张图片的速度增长。面对如此海量的图像数据，传统的图像存储和管理方法已经失效，如何根据图像内容自动地对大规模图像数据进行高效安全的存储、管理和检索，成为国计民生各领域的迫切需求。

[0009] 云计算作为一种崭新的计算和服务模式，采用资源租用、应用托管和服务外包的方式，为用户提供便捷的低成本计算和存储服务。

[0010] 云计算环境中，用户数据的存储与管理都外包给云服务提供商，用户无法按现有的安全模式来控制数据的访问与使用。用户数据的安全与隐私保护问题是阻碍云计算普及和推广的关键因素，特别是，Google, Apple, Salesforce.com等云服务提供商不断爆出各种安全事故，例如，2009年3月，Google发生大批用户文件外泄的事件，2010年6月，Apple iPad发生用户信息泄密事故，2011年12月，CSDN等网站上600万用户的密码被泄露，2013年苹果

iCloud云端照片的泄露事件等等,这些事件证实了人们对云安全的担心。如何在数据拥有者对其不具备直接物理控制的情况下,仍能保证数据的使用过程安全成为研究者们亟待解决的问题。

[0011] 为了保证用户图像数据的机密性,使企事业单位和个人将大规模隐私图像数据应用于云计算平台,放心地将自己的敏感图像数据交付给云服务提供商来管理,通常将敏感图像加密后再外包到云平台存储。但是,图像数据被加密后,数据之间的关联、冗余等原有特性消失了,现有明文图像的特征提取、加密方案、索引建立等大多都不支持对密文图像的运算。因此,研究云计算环境下的密文图像检索技术,即研究如何使图像拥有者能加密存储自己的隐私图像,又能让授权的图像使用者从海量图像中迅速找到自己需要的图像,是一个极具挑战性的课题。

[0012] 现有的密文数据检索的研究成果大多是关于文本检索,这些研究成果根据不同的安全模型,提出具有各种功能的检索方案,如可搜索加密、相似性搜索、多关键词排序搜索、动态搜索和索引构建等方案。相对而言,针对加密图像检索的研究方案较少。项等提出一种基于同态加密系统的图像可逆信息隐藏算法,该算法在保持数据量不变的前提下不仅完成同态加密域中额外信息的嵌入,而且降低了算法的复杂度、提高了信息的嵌入和提取效率。Yuan等提出一个全新的基于 SIFT和BOF无载体图像隐藏方案,该方案解决了不修改原始载体的秘密信息的隐藏的问题。Mishra等提出一种新的基于混沌映射的数字图像加密算法,该算法不但可以保证密钥的安全,而且能抵抗各种蛮力或统计攻击。Parvaz等人提出基于新的混沌系统的图像加密算法,该算法能够有效地抵抗差分,统计、噪声、数据等攻击,保证数据安全。

[0013] 以上几种密文图像方案取得较好的加密效果,对密文图像检索提供安全技术支持,但是加密图像检索不仅要保证数据的安全性,而且要考虑密文图像如何相似性检索。

[0014] Huang等人提出一种基于Henon映射的遥感图像可搜索加密方案。该方案将图像转换成特征向量,然后采用文本加密邻域中的相似度匹配算法来检索目标图像。该方案虽然有效地提高了检索加密遥感图像的安全性,但是不能进行高效的检索。Liu等提出一种基于云图像数据库的图像相似性增强隐私增强方案,虽然该方案提高了加密图像检索的安全性,但是该方案检索效率很低。Zhou等提出全局上下文验证方案来过滤错误匹配对来进行拷贝检测,在该方案中作者提出一种基于随机验证的快速图像相似度度量方法。虽然该算法有较高的检索精度,但是缺少对图像特征的处理,花费较多的检索时间。Hazra等人设计一个安全的加密检索系统。该系统采用HSV直方图作为图像特征,结合KNN与SVM算法查询相似图像,并取得较高的检索精度。

[0015] 以上几种加密图像检索方案可以保证图像的安全,也可以进行相似检索。但是这些方案都没有对图像构建索引,检索效率较低。因此,选择合理的索引构建算法是提高检索效率的关键。

[0016] Abdul jabbar等提出在物联网云下的加密检索方案,该方案采用SURF提取图像特征,采用LSH算法构建索引,并可以进行安全相似检索。该方案保证了智能端到云服务器端的数据安全,但是该方案没有对局部敏感算法优化。Xia等[1] 提出一个云环境下基于局部特征的加密图像检索方案,该方案采用SIFT提取图像特征,采用地球运动距离(EMD)评估图像的相似性,并采用局部敏感哈希算法构建哈希表。虽然该算法的检索效率有了提高,但是

该文章并没有对局部敏感哈希算法的参数优化。而且SIFT算法提取图像特征花费时间多。Xia等[2]提出一个云环境下基于隐私保护的图像检索方案,该方案通过引入局部敏感哈希算法和k近邻算法,有效提高搜索效率和保证数据的安全性。但是该方案没有对哈希函数和哈希表进行优化。

[0017] 因此,有必要设计一种云环境下的密文图像检索方法及系统。

发明内容

[0018] 本发明所要解决的技术问题是提供一种云环境下的密文图像检索方法及系统,该云环境下的密文图像检索方法及系统不仅缩短了特征提取时间,而且有效提高了加密图像检索效率。

[0019] 发明的技术解决方案如下:

[0020] 一种云环境下的密文图像检索方法,

[0021] (一)针对数据拥有者:

[0022] 对于一个图像集M,实施以下操作:图像集M,是用于从中检索出相似图形返回给查询用户;

[0023] (1)用 $Gen_{Harris}(Harris, M)$ 生成特征集为 $\{G_i\}_{i=1}^n$;

[0024] (2)用 $Gen_{feature}(\{G_i\}_{i=1}^n)$ 生成特征向量 $\{f_i\}_{i=1}^n$;

[0025] (3)用 $Build_{index}(\{f_i\}_{i=1}^n)$ 生成索引I,用 $Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 分别生成加密特征向量 $\{f'_i\}_{i=1}^n$ 、加密图像集M'、加密索引I';

[0026] (4)将 $\{f'_i\}_{i=1}^n$ 、M'、I'发送到云服务器,将加密密钥K发送给查询用户;

[0027] $Gen_{Harris}(Harris, M)$ 表示采用Harris算法对图像M提取图像特征;

[0028] $Gen_{feature}(\{G_i\}_{i=1}^n)$ 表示对特征集 $\{G_i\}_{i=1}^n$ 生成特征向量;

[0029] $Build_{index}(\{f_i\}_{i=1}^n)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 构建索引;

[0030] $Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 、图像集M、索引I进行加密;

[0031] (二)针对查询用户:

[0032] 对于查询图像集 M_q 执行以下操作:查询图像集 M_q 中可以是一幅图像,也可以是多幅图像,如果是多幅图像,查询的时候是一张一张的查,用户可以查询多张图像。

[0033] (1)用 $Gen_{Harris}(Harris, M_q)$ 生成特征集为 $\{G'_i\}_{i=1}^{n_q}$;

[0034] $Gen_{Harris}(Harris, M_q)$ 表示采用Harris算法对图像 M_q 提取图像特征;

[0035] (2)用 $Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$ 生成特征向量 $\{f_{qi}\}_{i=1}^{n_q}$;

[0036] $Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$ 表示对特征集 $\{G'_i\}_{i=1}^{n_q}$ 生成特征向量;

[0037] (3)用 $Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$ 生成陷门TD,然后将陷门TD发送到云服务器;

[0038] $Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$ 为产生陷门的函数;

[0039] 如果一个登陆处理系统允许一个特定的用户识别码,通过该识别码可以绕过通常

的口令检查,直观的理解就是可以通过一个特殊的用户名和密码登陆进行修改等操作。这种安全危险称为陷门,又称为非授权访问。陷门是在某个系统或某个文件中设置的“机关”,使得在提供特定的输入数据时,允许违反安全策略。例如,一个登录处理子系统允许处理一个特定的用户识别码,以绕过通常的口令检查。

[0040] (4) 用解密函数 $Dec_{data}(K, \mathcal{R})$ 解密返回相似图像 \mathcal{R} ;

[0041] (三) 针对云服务器:

[0042] 用 $Search(I, M', \{f'_i\}_{i=1}^n, TD)$ 算法进行检索并返回相似结果集 \mathcal{R} 。

[0043] $Gen_{Harris}(\cdot)$ 是指基于自适应阈值与Forstner方法进行Harris角点优选;

[0044] [这个步骤应该是当查询用户输入待查询的图像时,需要对待查询图像应用这个步骤的处理,图像拥有者在上传到云端之前需要将图像集进行这个过程的处理,即将特征提取完后,再加密上传到云端,也就是上传到云端的除了加密图像还有加密特征。

[0045] 确定候选节点后,再根据最大角点响应函数对候选点进行筛选,确定提取的预筛选特征点总数 c_1 ,最后结合Frostner算法确定最佳候选点总数 c_2 ,实现 Harris角点优选。

[0046] $Gen_{feature}(\cdot)$ 是指结合SURF算法描述Harris特征点,并结合词袋模型生成图像的特征向量。

[0047] $Build_{index}(\cdot)$ 是指构建哈希索引;采用基于 ρ 稳定的LSH函数族构建哈希表,作为哈希索引。

[0048] Harris角点优选包括以下步骤:

[0049] 步骤1:采用8邻域相似像素分析法确定候选集 C ;

[0050] 对于图像中的任一目标像素点 (x, y) ,计算该目标像素点与8邻域范围内像素点灰度差的绝对值 Δ ,通过与设定的阈值 t 相比较来确定是否相似,统计目标像素点与周围8个点的相似个数 $N(x, y)$,如下式所示:

$$[0051] \quad N(x, y) = \sum_{i, j} \chi(x + i, y + j) (-1 \leq i \leq 1, -1 \leq j \leq 1, \text{且 } i \neq 0, j \neq 0)$$

[0052] 式中:识别函数 $\chi(x + i, y + j) = \begin{cases} 1, & \Delta(x + i, y + j) \leq t \\ 0, & \text{otherwise} \end{cases}$,当 $2 \leq N(x, y) \leq 6$ 时,将该

目标像素点 (x, y) 视为候选点,用 C 表示候选点集合;

[0053] 步骤2:计算每个候选点的响应函数CRF (CRF的计算参见后文公式7),定义阈值 \mathcal{T} 为最大CRF值的 ρ 倍,即

$$[0054] \quad \mathcal{T} = \rho * CRF_{max}$$

[0055] 根据最大角点响应函数 \mathcal{T} 对候选点进行筛选 (CRF值大于 \mathcal{T} 的入选),确定提取的预筛选特征点总数 c_1 ,预筛选特征集 C_1 ;本发明取 ρ 为 0.01。

[0056] 步骤3:结合Frostner算法确定最佳候选点集 $G (G = \{(x_i, y_i)\}_{i=1}^{c_2})$ 以及最佳候选点总数 c_2 ;

[0057] 首先,以预筛选特征集 C_1 任意一点 (x_i, y_i) 为中心建立 $3*3$ 窗口,对该窗口内的每个点计算协方差矩阵 cov :

$$[0058] \quad \text{cov} = \begin{bmatrix} \sum J'_x{}^2 & \sum J'_x J'_y \\ \sum J'_x J'_y & \sum J'_y{}^2 \end{bmatrix};$$

[0059] 其中, J'_x , J'_y 是Robert梯度算子;

$$[0060] \quad J'_x = f(x+1, y+1) - f(x, y), \quad J'_y = f(x+1, y) - f(x, y+1);$$

[0061] f 是灰度函数, $f(x, y)$ 表示点 (x, y) 的灰度值;

[0062] 接着,计算特征点的权值 ω 和圆度 τ ;

$$[0063] \quad \omega = \frac{\det(\text{cov})}{\text{trace}(\text{cov})}$$

$$[0064] \quad \tau = \frac{4\det(\text{cov})}{(\text{trace}(\text{cov}))^2}$$

[0065] 其中 $\det(\text{cov})$ 是协方差矩阵 cov 的行列式, $\text{trace}(\text{cov})$ 是协方差矩阵 cov 的迹;然后,将 ω 、 τ 分别与给定阈值 $\mathcal{T}_\omega, \mathcal{T}_\tau$ 比较,将满足 $\omega > \mathcal{T}_\omega$ 且 $\tau > \mathcal{T}_\tau$ 的备选点加入特征集 C_2 ;然后在一定窗口(如以备选点为中心的 $5*5$ 窗口内)内,依据权值 ω 将满足条件 $\omega(x, y) = \max\{\omega(x, y)\}$ 的点加入最佳候选点集 G , 最佳候选点集 G 中的候选点的个数即为 $c_2 \cdot \max\{\omega(x, y)\}$ 最大值是指在一定窗口内特征点的取值, 取 ω 的最大值,

[0066] 生成图像的特征向量的步骤如下:

[0067] 步骤s1: 用 \mathcal{K} -mean 聚类算法对局部特征 $\{C_i\}_{i=1}^{c_2}$ 进行聚类, 形成一个视觉单词; $\{C_i\}_{i=1}^{c_2}$ 是 $G = \{(x_i, y_i)\}_{i=1}^{c_2}$ 在聚类中的表示;

[0068] 步骤S1a, 随机选择 \mathcal{K} 个点作为聚类中心 $\{C_1, C_2, \dots, C_{\mathcal{K}}\}$; 然后, 用下式计算特征集 G 中的每个数据点到这 \mathcal{K} 个聚类中心的距离 d , 并将数据点按距离分配到最近的聚类中心, 形成 \mathcal{K} 个簇 $\mathcal{U} = \{u_i\}_{i=1}^{\mathcal{K}}$;

$$[0069] \quad d = \sum_i^{c_2} \sum_j^{\mathcal{K}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

[0070] 其中, (x_i, y_i) 是 C_2 中的点, (x_j, y_j) 是 \mathcal{K} 个点作为聚类中心中的点;

[0071] 步骤S1b, 用下式计算簇的平均值 u , 指定这些值为新的聚类中心, 即视觉单词 $u = \{u_1, u_2, \dots, u_{\mathcal{K}}\}$;

$$[0072] \quad \bar{u}_j = \frac{1}{|u|} \sum_{j=1}^{\mathcal{K}} u_j$$

[0073] 其中, \bar{u}_j 表示第 j 个视觉单词 u_j 的平均值, 即 u_j 的总和是指簇中的数据点的特征值相加。与簇 u 的个数的比值, $|u|$ 是 u 的个数;

[0074] 步骤S1c: 重复以上步骤S1a和S1b, 直到聚类中心的值满足最小化均方误差MSE 函数收敛, 即本次与上次的差值小于预设值; 此时视觉单词表示为 $u' = \{u'_1, u'_2, \dots, u'_{\mathcal{K}}\}$; 每个 u' 是一个 \mathcal{K} 特征向量;

[0075] $MSE = \sum_{j=1}^{\mathcal{K}} \sum_{u_i \in u} |u_i - \bar{u}_j|^2$;

[0076] 步骤S2:得到视觉单词之后,将局部特征 C_i 按下式对应到视觉单词中;

[0077] $C_i = \omega' = \{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\}$;其中 $\{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\}$ 表示 C_i 对应到视觉单词 u' 的权重;

[0078] 统计整幅图像的视觉单词出现频率,视觉单词出现频率即统计每个视觉单词在局部特征出现的个数,生成图像特征向量 f_i ,所有图像的特征向量表示为 $\{f_i\}_{i=1}^n$ 。采用基于 ρ 稳定的LSH函数族构建哈希表作为哈希索引的步骤如下:

[0079] 数据拥有者选取 L 个LSH函数 $\{h_1, h_2, \dots, h_L\}$,并对所有的特征向量 $\{f_i\}_{i=1}^n$ 应用函数 $g(f_i) = (h_1(f_i), \dots, h_L(f_i))$; $g(f_i)$ 表示哈希函数族;为了提高准确率,重复这个过程 λ 遍,从而生成 λ 个哈希表;用 $\{D_{i,j}\}$, $i \in [1, \lambda]$, $j \in [1, N_i]$ 表示由局部敏感哈希函数生成的桶值集, N_i 表示第 i 个哈希表中的总桶数; $ID(m_i)$ 表示图像 m_i 将自身的ID关联至相应的桶值 $D_{i,j}$,形成加密哈希表。

[0080] 一种云环境下的密文图像检索系统,包括数据上传终端、云服务器和查询终端;

[0081] (1) 数据上传终端用于对图像拥有者将图像进行处理,将处理后的图像的加密特征上传到云服务器;

[0082] (2) 云服务器用于存储数据上传终端上传的图像以及图像的加密特征;云服务器还用于执行检索操作,将检索结果返回到查询终端;

[0083] (3) 查询终端用于查询用户输入待查询的图像,以及用于显示云服务器返回的查询结果;

[0084] 采用前述的云环境下的密文图像检索方法实施图像检索。

[0085] 本发明提出基于Harris角点优选与局部敏感哈希的加密图像检索,通过8邻域相似像素分析法与Forstner两方面对Harris角点进行筛选与优化,首先采用改进 Harris算法提取图像特征,接着采用SURF与词袋模型对图像特征聚类,生成每一幅图像的特征向量,然后采用局部敏感哈希算法构建索引,最后,采用文献[2] (文献2:Z.Xia,N.Xiong,V.Vasilakos,et al.EPCBIR:An efficient and privacy-preserving content-based image retrieval scheme in cloud computing[J]. Information Sciences,2017,387:195-204.)的加密方案加密数据,并采用欧式距离进行特征向量的相似度。

[0086] 有益效果:

[0087] 本发明提出的云环境下的密文图像检索方法及系统,是一种基于Harris角点优选与局部敏感哈希优化的加密图像检索方案。首先,从自适应阈值和特征点预筛选两方面优化Harris算法,并提取图像特征。其次,采用SURF (Speeded-Up Robust Features) 算法和词袋模型生成每一幅图像的Harris角点特征向量。然后,采用局部敏感哈希 (Locality Sensitive Hashing, LSH) 算法对特征向量构建可搜索索引,并用传统加密方案对图像以及索引进行加密,最后,在云服务器上进行安全的相似性检索。实验结果证明,通过对Harris角点优选及SURF与词袋模型的特征描述,并对局部敏感哈希算法的参数进行了优化,本发明提出的检索方案与现有加密检索方案对比,不仅缩短了特征提取时间,而且有效提高了加密图像检索效率。

附图说明

- [0088] 图1为加密检索模型示意图；
 [0089] 图2为生成特征向量的流程图；
 [0090] 图3为本发明方法与现有方法的检索精度对比图(之一)；
 [0091] 图4为本发明方法与现有方法的检索精度对比图(之二)；
 [0092] 图5为本发明方法与现有方法的搜索时间对比图；
 [0093] 图6为本发明方法与现有方法的特征提取时间对比图；
 [0094] 图7为本发明方法与现有方法的聚类时间对比图；
 [0095] 图8为本发明方法与现有方法的索引构建时间对比图；
 [0096] 图9为本发明方法与现有方法的陷门时间对比图。

具体实施方式

[0097] 以下将结合附图和具体实施例对本发明做进一步详细说明：

[0098] 实施例1:如图1,系统模型简介

[0099] 本发明采用文献[2]中类似的模型,如图1所示,具有三个模块:数据拥有者、授权用户和云服务商。

[0100] 云服务商提供基于内容的加密图像检索。授权用户生成并提交搜索请求给云服务商,由云服务商进行相似度比对,并将排序后的搜索结果返回授权用户。

[0101] 查询用户拥有数据拥有者的授权,对查询图像 $M_q = \{m_1, m_2, \dots, m_{n_q}\}$ m_1 表示图像库中某一幅图像,提取特征集 $\{G'_i\}_{i=1}^{n_q}$,并生成查询向量 $\{f_{qi}\}_{i=1}^{n_q}$, n_q 表示查询图像库 M_q 的图像数。特征集中任意一个 G'_i 可以表示一幅图像,每一个 G'_i 是[特征点数*64]维的特征向量,每一个查询向量表示一幅图像,是一个1*128维的向量,然后构建陷门TD,最后解密查询结果 \mathcal{R} 。

[0102] 数据拥有者要将图像集 $M = \{m_1, m_2, \dots, m_n\}$ (M 表示外包的图像库, M_q 表示查询图像库,通过提取查询图像库的特征,在外包的图像库检索相似图像)上传到云服务器,并保持搜索的能力, n 表示图像库 M 中图像 m 的数量。数据拥有者首先从图像集 M 中提取图像特征集合 $\{G_i\}_{i=1}^n$ 。接着生成特征向量 $\{f_i\}_{i=1}^n$ 并构建可搜索索引 I 。然后将 $\{f_i\}_{i=1}^n$ 、 M 和 I 加密发送到云端,为方便授权用户进行数据访问,最后数据拥有者需要将数据解密和搜索的一系列密钥信息发送给授权用户。相关知识介绍

[0103] 本发明采用改进的Harris算法提取图像特征,并用优化的局部敏感哈希算法对图像特征构建索引。在本节介绍Harris算法和局部敏感哈希算法。

[0104] 1Harris算法

[0105] Harris角点检测算法[3]是在Moravec算法基础上发展起来的,是由Harris C和Stephens M J提出的。Harris通过运用微分运算和自相关矩阵改进了Moravec角点检测算法。对于一幅图像 m ,以某像素点 (x, y) 为中心的小窗口在 x 方向上移动 u , y 方向上移动 v , Harris给出的灰度强度变化如公式(1)所示:

$$[0106] \quad E(x, y) = \sum w(x, y)[g(x+u, y+v) - g(x, y)]^2 \quad (1)$$

[0107] 其中, $g(x, y)$ 表示在 (x, y) 点的灰度值, g 表示灰度函数, $w(x, y)$ 为高斯滤波器, 如公式 (2) 所示。按照二阶泰勒级数展开灰度变化如公式 (3) :

$$[0108] \quad w(x, y) = \exp\left(\frac{x^2+y^2}{2\delta^2}\right), \delta \text{ 是参数, 一般取 } 1-2.$$

[0109] 中间的数, 本发明取 2; (2)

$$[0110] \quad E(x, y) = [u \quad v] \mathfrak{M} \begin{bmatrix} u \\ v \end{bmatrix} (3)$$

[0111] 二维矩阵 \mathfrak{M} , 表示如公式 (4)

$$[0112] \quad \mathfrak{M} = w(x, y) \otimes \begin{bmatrix} J_x^2 & J_x J_y \\ J_x J_y & J_y^2 \end{bmatrix} (4)$$

$$[0113] \quad J_x = g \otimes [-1, 0, 1] (5)$$

$$[0114] \quad J_y = g \otimes [-1, 0, 1]^T$$

(6)

[0115] 其中, J_x 表示 x 方向的梯度, J_y 表示 y 方向的梯度, 如公式 (5)、(6) 所示, \otimes 表示卷积, 这里卷积可以理解为 g 与矩阵的所有值相乘。

[0116] \mathfrak{M} 特征值的大小与特征点的性质相关。当两个特征值 λ_1, λ_2 均相对较大时, 根据 Harris 算法定义, Harris 算法的核心是利用局部窗口在图像上进行移动判断灰度值发生的变化相对较大指两特征值很多时, 此时灰度值也变化明显。此点即为需要提取的角点。角点的响应函数如公式 (7) :

$$[0117] \quad CRF = \det(\mathfrak{M}) - \Phi * (\text{trace}(\mathfrak{M}))^2 (7)$$

[0118] 其中, $\det(\mathfrak{M}) = \lambda_1 \lambda_2$, $\text{trace}(\mathfrak{M})$ 为矩阵 \mathfrak{M} 的迹, 即 $\text{trace}(\mathfrak{M}) = \lambda_1 + \lambda_2$, Φ 是经验参数, 本发明取 0.06, 当 CRF 取局部极大值且 CRF 大于设定的阈值时就是角点。

[0119] 2 局部敏感哈希算法

[0120] 局部敏感哈希 (Locality Sensitive Hashing) 是一种哈希函数族, 该函数族使得原始数据空间中的两个相邻数据点很大概率哈希到相同桶 (空间利用率较高的哈希表) 中, 根据这个特性可以用于近似查询 [4]。局部敏感哈希函数表示为: 由空间 \mathcal{S} 映射到集合 U 的哈希函数族 \mathcal{H} , 被认为是 (w, cw, p_1, p_2) 局部敏感的, 必须满足公式 (8) 的定义。如果对于任意两点 $x, y \in \mathcal{S}$, 存在

$$[0121] \quad \begin{cases} \Pr\{h(x) = h(y)\} \geq p_1 & d(x, y) \leq w \\ \Pr\{h(x) = h(y)\} \leq p_2 & d(x, y) \geq cw \end{cases} (8)$$

[0122] 其中, $d(x, y)$ 表示 x, y 的距离, 常数 $c > 1$, 概率 $p_1 > p_2$, w 是参数。可以通过增加哈希函数, 进而扩大 p_1, p_2 的差距, 达到提高效率的目的。

[0123] 基于 ρ 稳定的 LSH 函数族是局部敏感哈希函数的一种, 对于 ρ -LSH 函数 $h_{a,b}: \mathfrak{R}^l \rightarrow Z$, 该函数能将 1 维的向量 v 映射为一个整数 [4], 如公式 (9) 所示。

$$[0124] \quad h_{a,b}(v) = \left\lfloor \frac{a \cdot v + b}{w} \right\rfloor (9)$$

[0125] 其中a是一个服从高斯分布的1维向量;b∈[0,w]的实数。

[0126] *基于Harris角点优选与局部敏感哈希的加密图像检索方案

[0127] 为了减少数据拥有者前期的处理时间,并提高图像检索的效率。本发明首先对Harris算法优化,接着采用SURF算法结合词袋模型生成每副图像的特征向量,然后,采用优化的局部敏感哈希算法构建索引,最后,采用传统加密方案加密数据,并进行相似检索。

[0128] 1基于自适应阈值与Forstner的Harris角点优选

[0129] 基于内容的加密图像检索方案通常提取图像的局部特征,局部特征包括 SIFT特征、角点特征等。虽然SIFT特征有较好的鲁棒性,但是该算法特征提取时间多。与SIFT算法相比,Harris算法特征提取时间较少,因此为了减少特征提取时间本发明采用Harris算法提取特征。但是,Harris算法存在检测效率低、非极大值引起的伪角点多等问题,为了更好应用于加密图像方案,对Harris算法进行改进。首先,采用8邻域相似像素分析法确定候选集C。然后计算每个候选角点的响应函数CRF,根据最大角点响应函数对候选点进行筛选,确定提取的预筛选特征点总数 c_1 ,最后结合Frostner算法[5]确定最佳候选点总数 c_2 。

[0130] Step1采用8邻域相似像素分析法确定候选集C。具体流程:对于目标像素点(x,y),计算与8邻域范围内像素点灰度差的绝对值 Δ ,通过与设定的阈值 t 相比较来确定是否相似,统计目标像素点与周围8个点的相似个数 $N(x,y)$,如下式所示:

$$[0131] \quad N(x,y) = \sum_{i,j} \chi(x+i,y+j) (-1 \leq i \leq 1, -1 \leq j \leq 1, \text{且} i \neq 0, j \neq 0) \quad (10)$$

[0132] 式中: $\chi(x+i,y+j) = \begin{cases} 1, & \Delta(x+i,y+j) \leq t \\ 0, & \text{otherwise} \end{cases}$,由公式(1)知,当 $2 \leq N(x,y) \leq 6$ 时,将该点视为候选点,用C表示候选点集合。

[0133] Step2计算每个候选点的响应函数CRF,定义阈值 \mathcal{T} 为最大CRF值的 p 倍,如下式所示,根据最大角点响应函数 \mathcal{T} 对候选点进行筛选,确定提取的预筛选特征点总数 c_1 ,预筛选特征集 C_1 。基于自适应阈值的候选点的预筛选描述见算法1。

$$[0134] \quad \mathcal{T} = p * CRF_{max}(11)$$

算法 1: 基于自适应阈值的候选点的预筛选

输入: 图像 $m(x,y)$, 图像尺寸 $p \times q$, p , t

输出: 特征点集合 $C_2 = \{(x_i, y_i)\}_{i=1}^{c_1}$, $x < p, y < q$

1:初始化 $C \leftarrow \emptyset$, $C_1 \leftarrow \emptyset$, $c_1 = 0$
 2: for $x=1, \dots, p-8$ do
 [0135] 3: for $y=1, \dots, q-8$ do
 4: 根据公式(1)计算 $N(x,y)$
 5: if $2 \leq N(x,y) \leq 6$ then
 6: $C = C \cup (x,y)$
 7: end if
 8: end for
 9: end for

```

10: for x=1,...,p -8 do
11: for y=1,...,q -8 do
12: 根据(2)(4)(5)(6)(7)(11)计算 $\mathcal{T}$ 
13: if  $(x,y) \in C_1$ 且 $CRF(x,y) > \mathcal{T}$  then
[0136] 14:    $C_1 = C_1 \cup (x,y)$ ,  $c_1 = c_1 + 1$ 
15:   end if
16: end for
17: end for

18:输出特征点集合:  $C_1 = \{(x_i, y_i)\}_{i=1}^{c_1}$ 

```

[0137] Step3结合Frostner算法确定最佳候选点总数 c_2 。具体流程:首先,以预筛选特征集 C_1 任意一点 (x_i, y_i) 为中心建立 3×3 窗口,对该窗口内的每个点计算协方差矩阵 cov ,如公式(12)。

$$[0138] \quad cov = \begin{bmatrix} \sum J'_x{}^2 & \sum J'_x J'_y \\ \sum J'_x J'_y & \sum J'_y{}^2 \end{bmatrix} (12)$$

[0139] 其中, J'_x , J'_y 是Robert梯度算子。 $J'_x = f(x+1, y+1) - f(x, y)$, $J'_y = f(x+1, y) - f(x, y+1)$ 。 f 是灰度函数, $f(x+1, y+1)$ 表示该点的灰度值。接着,计算特征点的权值 ω 和圆度 τ ,如公式(13)、(14)。

$$[0140] \quad \omega = \frac{\det(cov)}{\text{trace}(cov)} (13)$$

$$[0141] \quad \tau = \frac{4\det(cov)}{(\text{trace}(cov))^2} (14)$$

[0142] 其中 $\det(cov)$ 是协方差矩阵 cov 的行列式, $\text{trace}(cov)$ 是协方差矩阵 cov 的迹。然后, ω 、 τ 与给定阈值 $\mathcal{T}_\omega, \mathcal{T}_\tau$ 比较,确定备选点特征集 C_2 。最后在窗口内,依据权值 ω 确定最佳候选点个数 c_2 。改进Harris与Forstner结合算法描述见算法2。

算法2: 改进 Harris 与 Forstner 结合算法

输入: 特征点集合 C_1 , 图像 $m(x, y)$, 图像尺寸 $p \times q$, \mathcal{T}_ω , \mathcal{T}_R

输出: 筛选后的特征点集合 $G = \{(x_i, y_i)\}_{i=1}^{c_2}$

```

[0143] 1: 初始化 $C_2 \leftarrow \emptyset$ ,  $G \leftarrow \emptyset$ ,  $\omega_{max} = 0$ ,  $c_2 = 0$ 
2: 计算 $J'_x$ ,  $J'_y$ , 结合公式(12)计算协方差矩阵 $cov$ 
3: for x=1: p -1 do
4:   for y=1: q -1 do

```

```

5: 结合 (13) (14) 计算特征点的权值 $\omega$ 、圆度 $\tau$ 
6: if  $\mathcal{T}_\omega < \omega$  &&  $\mathcal{T}_\tau < \tau$  then
7:    $C_2 = C_2 \cup (x, y)$ ,
8: end if
9: end for
10: end for
11: for i=1: p-1 do
12: for j=1: q -1 do
13: for x=1:5
[0144] 14: for y=1:5
15: if  $(x, y) \in C_2$  且  $\omega(x, y) = \max \{\omega(x, y)\}$  then
16:  $G = G \cup (x, y)$ ,  $c_2 = c_2 + 1$ 
17: end if
18: end for
19: end for
20: end for
21: end for

22: 输出筛选后的特征点集合:  $G = \{(x_i, y_i)\}_{i=1}^{c_2}$ 

```

[0145] 2结合SURF与词袋模型的特征描述

[0146] 为了提高Harris特征点的图像表征能力,提高图像检索精度与效率,本发明结合SURF算法[H.Bay,A.Ess,T.Tuytelaars,et al.Speeded-up robust features (SURF) [J].Computer Vision&Image Understanding,2008,110(3):346-359.]来描述Harris特征点,并结合词袋模型生成图像的特征向量。

[0147] 基于SURF的Harris特征描述算法如下:

[0148] Step1,构建Harris特征点尺度金字塔。

[0149] Step2,为Harris特征点选主方向,即G中每个特征点 (x_i, y_i) 为中心,以 $60s$ 为半径, s 是该点的尺度(尺度一般是指开展研究所采用的空间大小的量度),计算在 60° 扇形区域的特征点在水平和垂直方向的Haar小波响应总和(Haar小波的尺寸边长为 $4s$)记为 Σsum_i 。然后以一定间隔(一般间隔 45°)旋转 60° 扇形,选取 Σsum_i 最大值时扇形的方向作为该特征主方向。

[0150] Step3,将边长 $20s$ 的正方形,划分为16个小正方形窗口,正方形的选取是以特征点中心,用公式(15)计算每个窗口中的特征子向量 v ,这样每个小区域就有4个值,所以在:G特征集中,每个特征点 $\{C_i\}_{i=1}^{c_2}$ 就是 $16 \times 4 = 64$ 维的描述向量。

[0151] $v = (\Sigma d_x, \Sigma d_y, \Sigma |d_x|, \Sigma |d_y|)$ (15)

[0152] 其中, Σd_x 表示Haar小波特征的水平方向分量之和, Σd_y 表示Harr小波特征垂直方向分量之和, $\Sigma |d_x|$ 表示Harr小波特征的水平方向分量绝对值之和, $\Sigma |d_y|$ 表示 Haar垂直方向分量绝对值之和。

[0153] 用词袋模型生成每副图像的特征向量方法如下:

[0154] Step1用 \mathcal{K} -mean 聚类算法对局部特征 $\{C_i\}_{i=1}^{c_2}$ 进行聚类,形成一个视觉单词。

\mathcal{K} -mean 聚类流程:首先,随机选择 \mathcal{K} 个点作为聚类中心 $\{c_1, c_2, \dots, c_{\mathcal{K}}\}$ 。然后,用公式(16)计算特征集 G 中的每个数据点到这 \mathcal{K} 个中心的距离 d ,并将数据点按距离分配到最近的中心,形成 \mathcal{K} 个簇 $\mathcal{U} = \{u_i\}_{i=1}^{\mathcal{K}}$ 。

$$[0155] \quad d = \sum_i^{c_2} \sum_j^{\mathcal{K}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (16)$$

[0156] 其中, (x_i, y_i) 是 C_2 中的点, (x_j, y_j) 是 \mathcal{K} 个点作为聚类中心中的点;

[0157] 接着,用(17)公式计算簇的平均值 u ,指定这些值为新的聚类中心,即视觉单词 $u = \{u_1, u_2, \dots, u_{\mathcal{K}}\}$ 。

$$[0158] \quad \bar{u}_j = \frac{1}{|u|} \sum_{j=1}^{\mathcal{K}} u_j \quad (17)$$

[0159] 其中, \bar{u}_j 表示第 j 个视觉单词 u_j 的平均值,即根据公式(17)知,是 u_j 的总和与簇 u 的个数的比值, $|u|$ 是 u 的个数。

[0160] 最后重复以上步骤2和3,直到聚类中心的值满足最小化均方误差MSE函数式(18)收敛。此时视觉单词表示为 $u' = \{u'_1, u'_2, \dots, u'_{\mathcal{K}}\}$ 。

$$[0161] \quad MSE = \sum_{j=1}^{\mathcal{K}} \sum_{u_i \in u} |u_i - \bar{u}_j|^2 \quad (18)$$

[0162] Step2得到视觉单词之后,将局部特征 c_i 对应到视觉单词中,如公式(19)所示。统计整幅图像的视觉单词出现频率,生成图像特征向量 f_i ,所有图像的特征向量可以表示为 $\{f_i\}_{i=1}^n$ 。SURF与词袋模型的特征描述算法见算法3。

$$[0163] \quad c_i = \omega' = \{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\} \quad (19)$$

[0164] 其中 $\{\omega_1, \omega_2, \dots, \omega_{\mathcal{K}}\}$ 表示 c_i 对应到视觉单词 u' 的权重。

算法 3: SURF 与词袋模型的特征描述算法

输入: $G = \{(x_i, y_i)\}_{i=1}^{c_2}$, \mathcal{K} , 图像集 M

输出: f_i

1: 初始化 $u \leftarrow \emptyset, u' \leftarrow \emptyset, f_i \leftarrow \emptyset, c' \leftarrow \emptyset, sum = 0$

2: 构建尺度金字塔

[0165] 3: for $i=0, \dots, c_2$ do

4: 根据 Step2 计算 $\sum sum_i$, 即当选取 $\max\{\sum sum_i\}$ 时扇形的方向为特征的主方向。

5: 根据 Step3 $c' = c' \cup c_i$, 生成描述向量 $\{c_i\}_{i=1}^{c_2}$

6: end for

7: for $i=0, \dots, c_2$ do

8: 随机选择 \mathcal{K} 个数据对象作为聚类初始中心, $\{c_1, c_2, \dots, c_{\mathcal{K}}\}$

[0166]

9: 用公式(16)计算中心的距离 d , 形成 \mathcal{K} 个簇 $u \leftarrow u \cup u_i$
 10: for $j=1, \dots, \mathcal{K}$ do
 11: 用公式 (17) 计算每个聚类簇的平均值, 并更新聚类中心。 $u \leftarrow u \cup u_j$
 12: 根据公式 (18) 计算 MSE ,
 13: 重复 11-16 步骤, 直到 MSE 收敛, 此时视觉单词表示记为 $u' \leftarrow u' \cup u'_j$
 14: 根据词袋模型 Step2 生成图像的特征向量, $f_i = f_i \cup f_{i,j}$
 15: end for
 16: end for

[0167] 因为直接采用提取的特征向量进行相似性检索,不能取得高效检索效率。因此要选择合理的算法,达到高效检索的目的。

[0168] 3哈希索引构建

[0169] 为了提高检索效率,本发明采用一个预处理索引表对相似图像预筛选。现有的方法大多是采用局部敏感哈希算构建索引。本发明采用基于 ρ 稳定的LSH函数族构建哈希表, ρ -LSH具有局部敏感哈希算法的特性,因此本发明通过增加LSH 函数族的数量以及哈希表的数量,对该算法进行了优化。具体而言,数据拥有者选取 L 个LSH函数 $\{h_1, h_2, \dots, h_L\}$,并对所有的特征向量 $\{f_i\}_{i=1}^n$ 应用函数 $g(f_i) = (h_1(f_i), \dots, h_L(f_i))$ 。为了提高准确率,重复这个过程 λ 遍,从而生成 λ 个哈希表。用 $\{D_{i,j}\}$,其中, $i \in [1, \lambda]$ $j \in [1, N_i]$ 。表示由局部敏感哈希函数生成的桶值集, N_i 表示第 i 个哈希表中的总桶数。 $ID(m_i)$ 表示图像 m_i 将自身的ID关联至相应的桶值 $D_{i,j}$ 如表1所示。为提高安全性,本发明运用一个单向函数对哈希表中的关键词进行加密。

[0170] 表1第 i 加密哈希表

[0171] Tab1Encrypted hash table of i

$\phi: (k_i, D_{i,1})$	$ID(m_1)$	、
	$ID(m_4)$ 、 $ID(m_7)$ 、 $ID(m_{11})$...	
$\phi: (k_i, D_{i,2})$	$ID(m_9)$	、
	$ID(m_{12})$ 、 $ID(m_{15})$ 、 $ID(m_{18})$...	
...	...	
$\phi: (k_i, D_{i,N_i})$	$ID(m_{13})$	、
	$ID(m_{16})$ 、 $ID(m_{19})$ 、 $ID(m_{22})$...	

[0173] 哈希索引表可以将相似的图像保存到同一个桶中,用户在搜索时可以缩短查询时间,但是这样搜索的结果仍然会有不相似图像,不能满足要求。因此,为了进一步提高检索精度,本发明用欧式距离对已经搜索的相似图像进行相似度量。综上所述,本发明需要设计一个特征提取时间少,并可以进行安全的检索的密文图像检索方案。

[0174] 4基于Harris角点优选与局部敏感哈希的密文图像检索方案

[0175] 本发明的加密图像检索方案有三个模块组成:数据拥有者、查询用户、云服务商。三个模块都有各自的任务,一起组成整个加密图像检索系统。

[0176] 数据拥有者:对于一个图像集 M ,首先,用 $Gen_{Harris}(Harris, M)$ 生成特征集为 $\{G_i\}_{i=1}^n$,

[0177] 接着用 $Gen_{feature}(\{G_i\}_{i=1}^n)$ 生成特征向量 $\{f_i\}_{i=1}^n$ 。然后用 $Build_{index}(\{f_i\}_{i=1}^n)$ 生成索引 I ,用 $Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 分别生成加密特征向量 $\{f'_i\}_{i=1}^n$ 、加密图像集 M' 、加密索引 I' 。最后,将 $\{f'_i\}_{i=1}^n, M', I$ 发送到云服务器,将加密密钥 K 发送给查询用户。 $Gen_{Harris}(Harris, M)$ 表示采用Harris算法对图像 M 提取图像特征; $Gen_{feature}(\{G_i\}_{i=1}^n)$ 表示对特征集 $\{G_i\}_{i=1}^n$ 生成特征向量; $Build_{index}(\{f_i\}_{i=1}^n)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 构建索引; $Enc_{data}(\{f_i\}_{i=1}^n, M, I)$ 表示对特征向量 $\{f_i\}_{i=1}^n$ 、图像集 M 、索引 I 进行加密,加密特征向量和索引的方法采用Xia[2]的方案,用Xia[2]的方案生成的密钥为 K 。

[0178] 查询用户:对于查询图像集 M_q ,首先,用 $Gen_{Harris}(Harris, M_q)$ 生成特征集为 $\{G'_i\}_{i=1}^{n_q}$,接着用 $Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$ 生成特征向量 $\{f_{qi}\}_{i=1}^{n_q}$ 。然后,用 $Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$ 生成陷门 TD ,然后将陷门 TD 发送到云服务器。最后,用 $Dec_{data}(K, \mathcal{R})$ 解密返回相似图像 \mathcal{R} 。

[0179] 云服务器:用 $Search(I, M', \{f'_i\}_{i=1}^n, TD)$ 算法进行检索并返回相似结果集 \mathcal{R} 。具体算法流程算法4。

算法 4 加密检索方案流程

[0180]	<p>数据拥有者</p> <p>Step1: $\{G_i\}_{i=1}^n \leftarrow Gen_{Harris}(M)$, 采用改进的 Harris 算法对图像集$M$提取特征$\{G_i\}_{i=1}^n$。</p> <p>Step2: $\{f_i\}_{i=1}^n \leftarrow Gen_{feature}(\{G_i\}_{i=1}^n)$, 结合SURF, BOW算法生成特征向量$\{f_i\}_{i=1}^n$。</p> <p>Step3: $I \leftarrow Build_{index}(\{f_i\}_{i=1}^n)$, 采用优化的$\rho - LSH$算法构建索引。</p> <p>Step4: $I', M', \{f'_i\}_{i=1}^n \leftarrow Enc_{data}(\{f_i\}_{i=1}^n, M, I)$, 本发明用混沌加密算法对图像集$M$加密, 加密图像为$M'$, 加密特征向量和加密索引分别是$I', \{f'_i\}_{i=1}^n$。</p> <p>查询用户</p> <p>Step1: $\{G'_i\}_{i=1}^{n_q} \leftarrow Gen_{Harris}(M_q)$, 采用改进的 Harris 算法对图像集$M_q$提取特征$\{G'_i\}_{i=1}^{n_q}$。</p> <p>Step2: $\{f_{qi}\}_{i=1}^{n_q} \leftarrow Gen_{feature}(\{G'_i\}_{i=1}^{n_q})$, 结合SURF, BOW算法生成查询向量$\{f_{qi}\}_{i=1}^{n_q}$。</p>
--------	--

[0181]

Step3: $TD \leftarrow Gen_{TD}(K, \{f_{qi}\}_{i=1}^{n_q})$, 用密钥 K 加密查询向量生成陷门 TD

Step4: $\mathcal{R} \leftarrow Dec_{data}(K, \mathcal{R})$, 用密钥解密返回的图像集 \mathcal{R}
云服务器

Step1: $\mathcal{R} \leftarrow Search(I, M', \{f'_i\}_{i=1}^n, TD)$ 在云服务器进行安全的相似性检索, 并返回查询用户结果集 \mathcal{R}

[0182] 实验结果与分析

[0183] 本发明用Matlab R2014a+vs2008C++在Dell-14R-5421笔记本电脑、操作系统Windows 10系统、CPU为Intel (R) Core (TM) i5-3337U@1.80GHz对算法测试。本发明实验的图像库是Corel test set[7]。图像库中分为10类, 每一类100幅大小 256×384 或 384×256 的JPG格式的图像。本发明实验在 Harris算法、 L 、 λ 三个方面有效优化加密图像检索。实验中使用的参数: $w=4$, $\mathcal{K} = 128$, 每个LSH函数使用的函数用 L 表示。

[0184] 1检索精度

[0185] 本发明中的搜索精度可以定义为: $P_k = k' / k$, 其中 k' 是返回的前 k 个图像中与查询图像相似的图像。本发明的实验思路: 首先, 从10类图像随机抽取两幅图像组成查询库, 然后测试当 $k=10, 20, 25, 30, 35, 40, 45, 50$ 时, 本发明与Xia等的检索方案的检索精度。

[0186] 图3-4是当 L 、 λ 取不同参数时, 根据Top-k的取值, 检索效率变化结果图。当 $L=2$ 、 $\lambda=20$, 且Top-k小于25时, 本发明的检索效率与Xia的方案检索效率持平, 检索效率随着Top-k的增大而减小。但当Top-k大于25时本发明的检索效率略优于Xia的方案, 如图3。调整 L 、 λ 的值, 当 $L=2$ 、 $\lambda=2$, 此时两种方案与原来相比, 检索效率有所降低, 但本发明算法的检索效率仍高于Xia的方案, 如图4。通过这个实验结果分析, 本发明的检索效率比Xia的方案高效, 本发明的搜索时间与Xia的方案较好。

[0187] 2检索时间

[0188] 本发明从两个方面衡量检索时间, 第一, 索引构建时间。第二, 陷门生成时间。

[0189] 索引构建

[0190] 在采用 ρ -stable局部敏感哈希算法构建索引前, 首先应提取图像特征, 然后采用词袋模型聚类生成特征向量。因此本索引构建的时间可以分为三个部分: 第一特征提取时间, 第二聚类时间, 第三索引构建时间。

[0191] 图6, 7分别是两个方案的特征提取时间和聚类时间。从图中可以看出, 本发明的检索方案不但特征提取时间比Xia的特征提取时间少, 而且聚类的时间也取得较好结果。图7是索引构建时间的对比。本发明取与Xia的文献[1]相同的参数进行测试比较, 即 $L=2$ 、 $\lambda=20$, 本发明的索引构建时间低于Xia的方案, 如图8。从整个索引构建时间分析, 本发明比Xia的方案高效。

[0192] 陷门构建

[0193] 陷门构建时间如图9所示。分析整个陷门时间变化, 与Xia的方案相比, 本发明构建陷门的时间较短。

[0194] 结论

[0195] 本发明提出一个云环境下基于Harris角点优选与局部敏感哈希优化的加密图像检索方案。

[0196] 该方案首先优化了Harris算法固定阈值和运算速度慢的问题,并结合Forstner算法提取图像特征,接着,为了更好的在云服务器上检索,采用SURF算法描述每个特征点,并结合词袋模型生成每副图像的特征向量。然后,为了提高检索效率,优化局部敏感哈希算法参数,并用该算法构建索引。最后,在云服务器上对加密的数据进行相似性检索。安全分析和实验表明了本方案的安全性和高效性。

[0197] 文献1和文献2出处:

[0198] [1]Z.Xia,Y.Zhu,X.Sun,et al.Towards Privacy-preserving content-based image retrieval in cloud computing[J].IEEE Transactions on Cloud Computing, 2018,6(1): 276-286.

[0199] [2]Z.Xia,N.Xiong,V.Vasilakos,et al.EPCBIR:An efficient and privacy-preserving content-based image retrieval scheme in cloud computing[J].Information Sciences, 2017,387:195-204.

[0200] [3]C.Harris,M.Stephens.A combined corner and edge detector[J].Proc Alvey Vision Conf.IEEE, 1988(3).147-151.

[0201] [4]M.Datar,N.Immorlica,P.Indyk,V.S.Mirroknj.Locality-sensitive hashing scheme based on p-stable distributions[C]//Twentieth Symposium on Computational Geometry.ACM, 2004:253-262.

[0202] [5]W.Förstner,E.Gülch.A fast operator for detection and precise location of distinct points, corners and circular features[J].Isprs Intercommission Workshop Interlaken,1987:281-305.

[0203] [6]H.Bay,A.Ess,T.Tuytelaars,et al.Speeded-up robust features (SURF) [J].Computer Vision& Image Understanding,2008,110(3):346-359.

[0204] [7]J.Wang,J.Li,G.Wiederhold.Simplicity:Semantics-sensitive integrated matching for picture libraries[J].Pattern Analysis&Machine Intelligence IEEE Transactions on,2001,23(9):947-963.

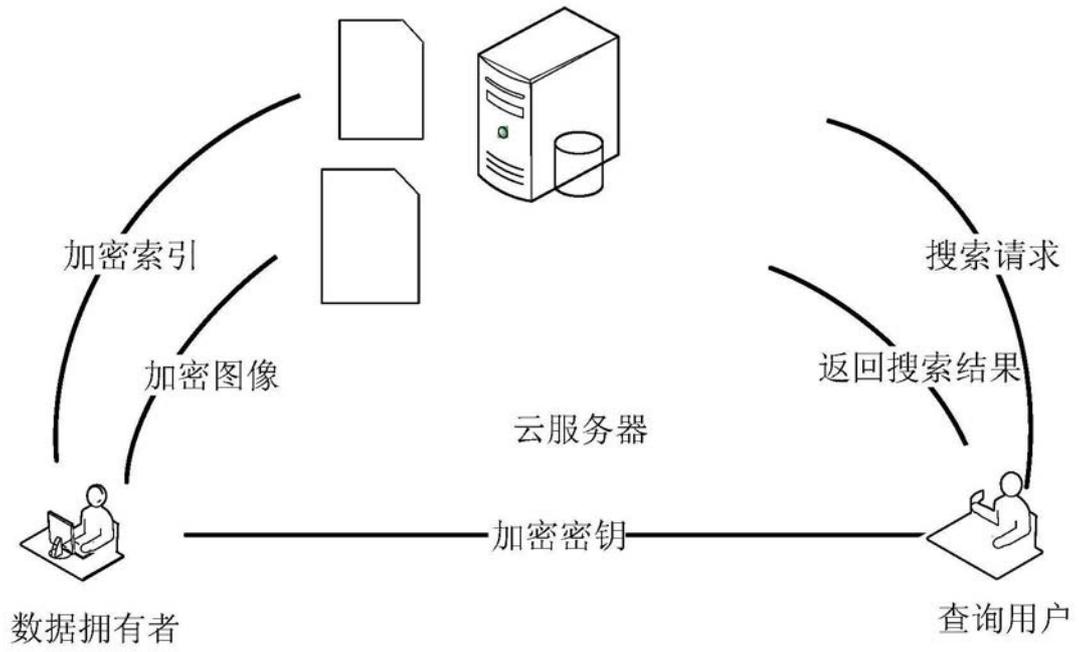


图1

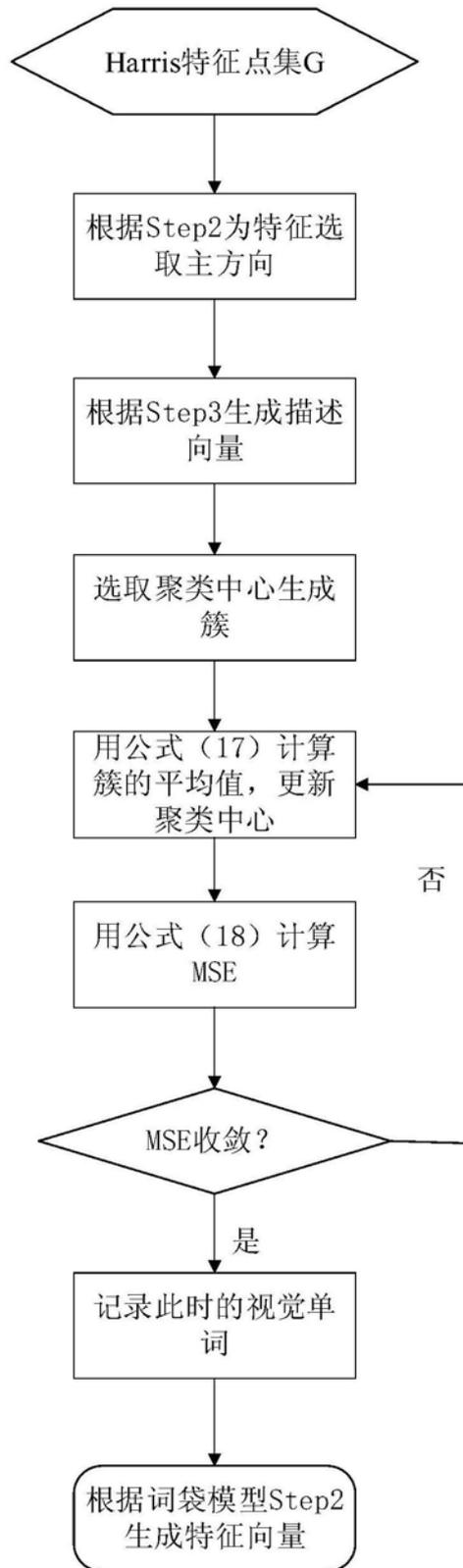


图2

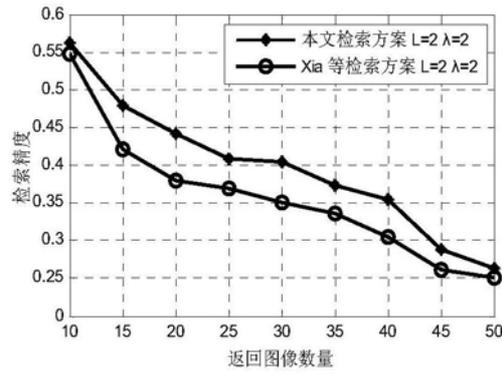


图3

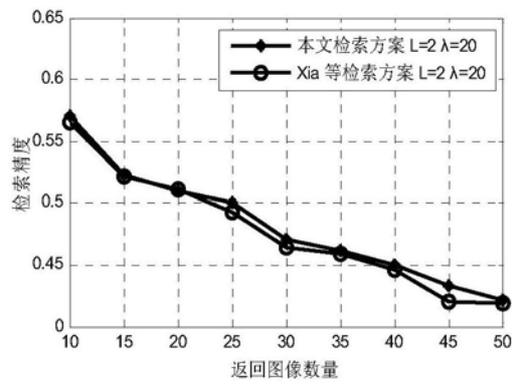


图4

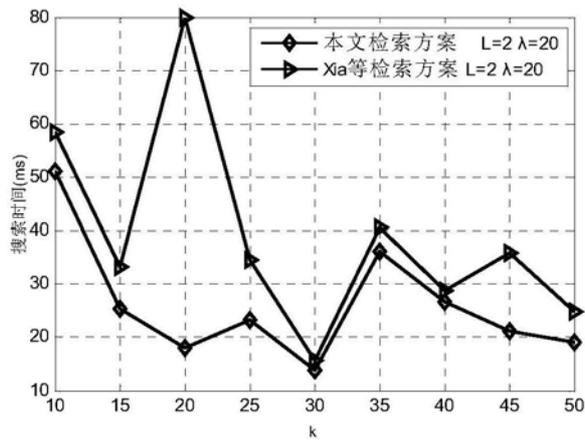


图5

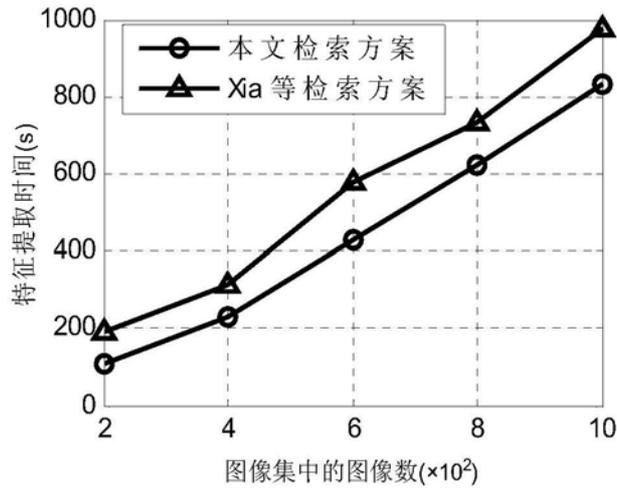


图6

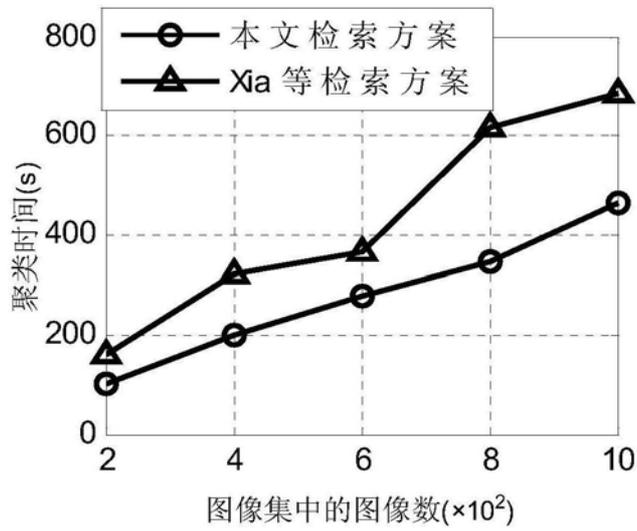


图7

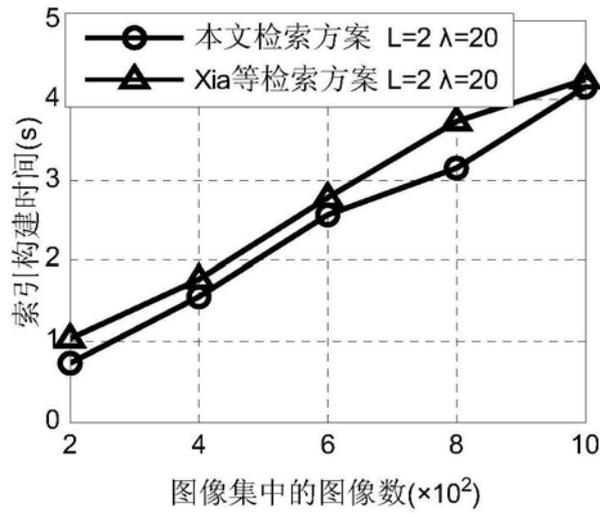


图8

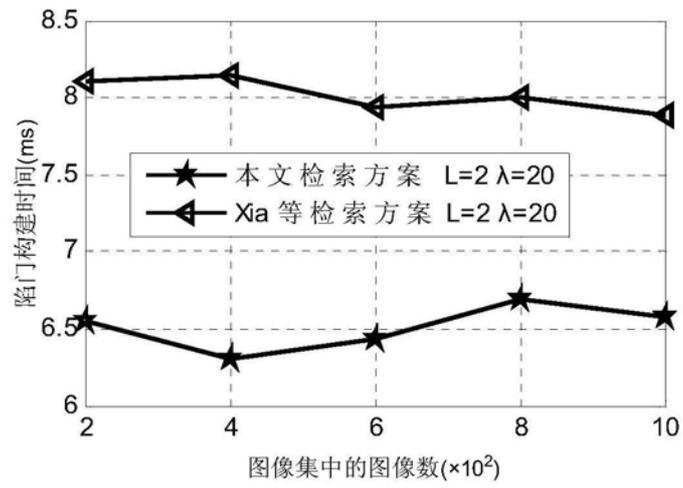


图9