ImageProof: Enabling Authentication for Large-Scale Image Retrieval

2021 年 9 月 6 日

1 摘要

随着在线图像的爆炸式增长和搜索引擎的普及，中小企业建立大规模的图像检索系统给云外包平台产生了巨大的需求。在减少存储和检索负担时，企业也面临着不受信任的云服务提供商的风险。

由于图像文件规模很大，主要挑战是：

1. 设计有效的身份验证数据结构（ADS）

2. 平衡搜索、通信和验证的复杂性

为了解决这些问题，提出两种新的 ADS：Merkle 随机 k-d 树和 Merkle 反向索引，确保了查询结果的完整性。

最终提出了名为 ImageProof 的图像检索认证方案。

2 引言

基于内容的图像检索（Content-based image retrieval）CBIR 是一个从大型图像数据库中搜索具有相似内容的图像系统。在多种现成的 CBIR 方案中，尺度不变特征变化（scale invariant feature transform）SIFT 是一种广泛使用的检测和提取图像局部特征的方法。基于此方法能使邻近索引算法能够识别相似图像，已经在许多 CBIR 系统中实现。

现在的图像安全检索工作主要集中在诚实但好奇的模型上。假定服务提供商 SP 可以返回真正的搜索结果，但是现实情况却不仅于此。

查询身份验证技术保证了外包数据库搜索结果的完整性，是一种防止不信任的 SP 的方案。其基本思想是，数据所有者设计一个 ADS 将其与原

始数据库外包给 SP。收到查询请求后，SP 根据 ADS 计算结果以及验证对象 VO 的加密证明。搜索结果和 VO 都将返回给客户端。使用 VO，客户端可以验证接受到的搜索结果是否真实。

尽管查询验证已经得到了广泛的研究，但是没有用于大规模图像检索验证技术。问题有两方面：一方面，CBIR 是一个大型的、复杂的系统。尤其是 SIFT 图像检索相似图像时通常需要两个步骤，袋视觉单词 BoVW 编码和反向索引搜索。

BoVW 编码可以将查询图像转换为稀疏的 BoVW 向量，而反向索引可以方便使用这些向量进行图像搜索，但是现在的查询验证方案都不能同时实现经过身份验证和 BoVW 编码和反向索引搜索。

另一方面，客户端有限的存储、通信和计算资源。如果只外包反向索引，那么客户端的计算和通信成本会很高。

3 背景知识和准备工作

3.1 基于 SIFT 的图像检索

目标是搜索顶部 k 个相似的图像。

一般过程是两个步骤：BoVW 编码和反向索引搜索。

采用了一种常用的技术，近似 k-means（AKM）用于 BoVW 编码

图像由提取出的一组特征向量表示。大量的聚类也被称为代码本或者词汇表，根据数据库中图像的特征向量进行预先训练。

给定一个图像，使用 AKM 寻址图像的每个特征向量的近似最近聚类。通过计算这些近似的聚类频率，得到图像的稀疏 BoVW 向量，第 i 个值表示近似最接近第 i 个聚类的特征向量的数量。

假设 Q 为查询图像，I 为数据库中的图像，B_Q、B_I 是 Q、I 的 BoVW 向量。Q 和 I 之间的相似性由 B_Q 和 B_I 之间的相似性定义。

3.1.1 相似度测量

符号	描述
w_{c_i}	预训练的聚类 c_i 的权重
n_C	数据库中聚类的数量
n_D	数据库中的图像数量
n_{D,c_i}	至少有一个特征向量与聚类 c_i 的图像数量
f_{I,c_i}	B_I 中聚类 c_i 的频率
p_{I,c_i}	I 中聚类 c_i 的影响

则定义如下

$$w_{c_i} = ln\frac{n_D}{n_{D,c_i}} \tag{1}$$

$$p_{I,c_i} = \frac{w_{c_i} \cdot f_{I,c_i}}{||B_I||} \tag{2}$$

其中 $||B_I||$ 是 B_I 的 $L_2 - norm$
则 Q 和 I 的相似性根据余弦距离定义：

$$S(Q,I) = \sum_{i=1}^{n_c} p_{Q,c_i} \cdot p_{I,c_i} \tag{3}$$

由于 BoVW 向量是稀疏的，大多数的值为 0，因此相似性可以简化为

$$S(Q,I) = \sum_{p_{Q,c_i} \neq 0} p_{Q,c_i} \cdot p_{I,c_i} \tag{4}$$

3.1.2 定义 1：顶级 k 图像搜索

给定一个访问的图片 Q，找到 k 个最相似的图像 $\{I_{top_i}\}_{i=1}^{k}$。
对于 $\forall i \in [1,k]$, 以及数据库中任意图像 $I \notin \{I_{top_i}\}_{i=1}^{k}$ 都有

$$S(Q, I_{top_i}) \geqq S(Q,I)$$

3.1.3 估计 k-means，AKM

给定一个特征向量，AKM 不是在大量预先训练的簇中找到精确的最近的簇，而是使用随机 k-d 树获得一个近似的最近的簇。

首先在预先训练的簇上构建一组随机的 k-d 树。与常规的 k-d 树相比，在索引树的每个层次上，随机的 k-d 树从方差最大的维数中随机选择分割维数。在处理一个给定的特征向量时，所有的树都会被遍历，直到叶子节点，该全局优先队列维护了特征向量到每个索引子空间的距离，在达到预定义的树遍历数后，将终止搜索，并返回最接近的聚类作为结果。

3.1.4 反向索引

主要由聚类和它们的反向列表组成，反向列表包含的所有图像的特征向量至少含有一簇与 c_i 相近。由一系列的 $< I, value >$ 委派表示，其中 I 是一个图像标识符，$value$ 是相关权重。

3.2 密码学原语和准备工作

3.2.1 数字签名

验证消息真实性和完整性的数字签名方案。

通过三种算法组成:

1. 私钥和公钥的秘钥生成算法 2. 给定消息和私密钥计算签名的签名算法 3. 以消息、签名和公钥作为输入和输出接收或拒绝的验证算法

3.2.2 Merkle Hash 树

一个加密哈希函数，表示为 $h(\cdot)$，以一个任意长的字符串作为输入，输出一个固定长度的位字符串。是一个防碰撞的单向函数，Merkle 哈希函数（MH-tree）是一个经过身份验证的二进制树，使用户在不检索整个数据库的情况下验证单个数据对象。

3.2.3 Cuckoo 过滤器

Cuckoo 过滤器是一种支持近似集合关系测试的数据结构，是 Cuckoo hash 表的一个紧凑变体，使用小的 f 位指纹来代表数据。由两个哈希函数 $h_1(\cdot)$ 和 $h_2(\cdot)$ 组成的桶数组，决定了一个输入项两个可用桶

4 问题设定

4.1 系统总览

三部分组成：图像拥有者，服务提供者 SP 和客户端。

在系统建立时，图像拥有者首先使用现有的特征提取法从图像中提取特征向量，构建经过 ADS

然后将图像数据集和 ADS 外包给 SP

之后，当 SP 接收到包含一组特征向量的查询请求时，SP 执行 BoVW 编码，根据定义 1，搜索顶部 k 相似图像。

SP 在 ADS 的基础上搜索结果构造 VO，将结果和 VO 返回给客户端

4.2 威胁模型

假定一个恶意模型，SP 可能会返回错误的搜索结果（伪造或是排名低的图像）

为了保护客户端不接收到错误的图像结果，客户使用 SP 返回的 VO 验证结果的完整性，结果会满足如下的安全属性：

1. 可靠性：结果必须是外包且未被修改的图像

2. 完整性：结果必须包括 k 个最相似的图像，即其他图像的相似性小于返回的图像

因此，问题就是如何设计 ADS，支持有效的查询处理和结果验证。

5 验证随机 k-d 树和带 Cuckoo 过滤器的反向索引

5.1 Merkle 随机 k-d 树

5.1.1 数据结构

MRKD 树由两种类型的节点组成：内部节点和叶子节点，内部节点有三个组件：分裂平面、指向子节点的指针和 digest。

内部结点的 Digest

l_{N_i}