文章编号：2095-2163(2019)06-0089-04 中图分类号：TP391 文献标志码：A

面向云计算的隐私保护图像特征提取方法研究

张晓璐

(福建林业职业技术学院 自动化工程系，福建 南平 353000)

摘 要：本文提出一种有效、实用的隐私保护方法，用于云计算环境中大规模加密图像数据的特征提取。为了保证运算的安全性，设计了安全乘法协议(SMP)和安全比较协议(SCP)。该方法对原始图像数据进行随机分割，并将特征提取操作分配到两个不同的云服务器上，以保持其关键特性，同时实现效率和安全性要求。实验结果表明，与现有的方法相比，该方法具有较高的召回率和精确度，并具有较低的计算和通信开销。

关键词：云计算；特征提取；隐私保护

Research on image feature extraction method for privacy protection for cloud computing

ZHANG Xiaolu

(Automation Engineering Department, Fujian Forestry Vocational Technical College, Fujian Nanping 353000 China)

【Abstract】 This paper proposes an effective and practical privacy protection method for feature extraction of large-scale encrypted image data in a cloud computing environment. In order to ensure the security of the operation, this paper designs the Secure Multiplication Protocol (SMP) and the Security Comparison Protocol (SCP). The method randomly segments the original image data and distributes the feature extraction operations to two different cloud servers to maintain its key characteristics while achieving efficiency and security requirements. The experimental results show that compared with the existing methods, the method has higher recall rate and accuracy, and has lower computation and communication overhead.

【Key words】 Cloud computing; feature extraction; privacy protection

0 引 言

机器学习技术可以从大规模多媒体数据中发掘有价值的知识和隐藏的信息，但在本地处理大量多媒体数据是一项耗时的工作。由于云计算的快速发展和云计算服务的普及，数据所有者利用丰富的云计算资源，将大量的图像数据和图像处理任务迁移到远程的云服务器上，以节省成本和增加服务的灵活性。然而，由于数据所有者和云所属的信任域不同，云计算环境中的数据存储和计算也引发了很大的安全性和隐私问题[1-2]。作为机器学习中的关键预处理步骤，研究者在保护隐私的云计算框架中对特征提取进行了广泛的研究，以便有效地去除不相关和冗余的数据，并提高学习准确性。在现有文献中，研究者提出了各种隐私保护计算，包括模幂运算、线性方程和 kNN 搜索。这些工作主要关注数值数据或文本数据的工程计算问题。近几年，密文域中的隐私保护数据搜索已经扩展到基于内容的多媒体检索、人脸识别和指纹识别，以及如何在云计算环境中实现安全图像搜索。然而，现有的研究均假设图像已经由特征提取算法预处理以获得图像的矢量表示。由于图像特征提取在多媒体数据处理中的重要性及其对海量数据的繁重操作，特别是对于其巨大尺寸和大量特征点的卫星数据，从密文域提取或检测图像特征已开始吸引越来越多研究人员的研究兴趣。

Hsu 等人[3]采用同态加密、探讨加密域中隐私保护尺度不变特征变换(SIFT)；然而该解决方案具有很高的计算开销。Qin 等人[4]借助于保持加密和随机排列，提出了一种改进的方案。Wang 等人[5]考虑基于形状特征提取的安全和私有外包问题，并分别通过使用同态加密和乱码电路协议[6]提出了两种具有不同安全级别的方法。虽然在隐私或效率方面付出了巨大努力，但是先前解决方案的一个共同限制是都缺乏关于保留原始图像特征提取算法的关键特性的全面分析和评估。Hu 等人[7]借助于乱码电路来解决这个问题，尽管该解决方案很好地保留了 SIFT 的关键特性，但仍有进一步降低其计算和通信成本的空间。更重要的是，其仍然无法消除边缘响应，使得检测到的关键点对于少量噪声不稳定。因此，如何在海量图像数据上实现隐私保护图像特

作者简介：张晓璐(1984-)，女，硕士，副教授，主要研究方向：计算机图形图像。

收稿日期：2019-09-10

征提取,同时减轻计算和通信负担是一个具有挑战性的问题。

1 问题描述

考虑一个基于云的多方图像特征提取计算模型:数据所有者 O、云服务器 $S1$ 以及云服务器 $S2$。假设数据拥有者 O 持有大量敏感图像文件,而数据所有者缺少图像处理所需的资源。因此, O 会将图像处理任务分配到具有丰富存储和计算资源的云。假设 $S1$ 和 $S2$ 分别属于两个不同的云服务提供商。为了保护自身的隐私, O 首先对每个图像集进行加密处理,然后将密文(即加密后的图像数据)分发到 $S1$ 和 $S2$。通过一系列安全交互协议在密文域中运行 SIFT 算法之后, $S1$ 和 $S2$ 将加密的特征描述符返回给数据所有者,数据所有者可以从其加密版本中恢复真实特征描述符。

现有研究指出,将图像特征提取任务分配给一个云实体容易导致隐私的泄漏。因此,有必要使用至少两个云实体来实现隐私保护。与此同时,假设云实体是半可信的[8]。

本文的目标是实现具有隐私保护的图像特征提取,同时尽可能保留图像的关键特征。为此,该方法需要实现以下的设计目标:

(1)安全约束。原始图像内容的隐私不能被泄露;

(2)有效性约束。由该方法提取的特征向量应尽可能接近原始 SIFT 算法的结果;

(3)效率约束。与单独执行 SIFT 算法相比,数据所有者仅需要进行少量的计算。

2 安全运算协议

2.1 安全乘法协议 SMP

假设两个服务器 S_1 和 S_2 分别拥有私有的 k 比特长度的整数 x_1 和 y_1 的。在执行协议之后, S_1 能够接收到一个服从均匀分布的随机值 s_1, 而 S_2 接收到另一个服从均匀分布的随机值 r_1, 使得 $s_1 + r_1 = x_1 \times y_1 (\mathrm{mod}\ p)$。

本文的安全乘法协议(SMP)能使双方安全地计算多对私有整数的乘积,使计算和通信成本大大降低。由于仅考虑 8 位长度的输入和 16 位的 p 值,乘积中间结果的长度会远小于 p 的长度。因此,为了便于说明,在下文中将取模操作省略。在 SMP 开始时, S_1 将首先使用 SIMD 以打包方式加密 x, 然后将加密后的 x 发送给 S_2。S_2 将利用同态属性将其与 y 相乘。最后, S_2 将使用随机生成的数字加密乘法操作的结果,并将其发送到 S_1 进行解密。

2.2 安全比较协议

安全比较协议(SCP),能够在隐私得到保护的条件下将两个整数进行比较。SCP 协议通过结合 SHE 与 SIMD 来优化安全标量积协议(SPP),以实现在对多对整数进行比较的同时保护隐私,并使通信和计算开销减少。

假设有两个 k 位的整数 $x^k x^{k-1} \ldots x^1$ 和 $y^k y^{k-1} \ldots y^1$, 对于整数中的第 i 位,有:

$$\begin{cases} a^i = x^i \times (1 - y^i), b^i = y^i \times (1 - x^i). \\ \qquad i = k \\ a^i = (1 - b^{i+1}) \times (a^{i+1} + (1 - a^{i+1}) \times x^i \times \\ \qquad (1 - y^i)), \\ b^i = (1 - a^{i+1}) \times (b^{i+1} + (1 - a^{i+1}) \times y^i \times \\ \qquad (1 - x^i)). \\ \qquad i < k \end{cases} \tag{1}$$

如果 a^1(b^1)等于 1,则认为 x 比 y 大(小)。S_1 将首先对 x 进行逐位加密并将密文发送到 S_2。然后 S_2 将基于同态性质在密文上采用公式(1)进行计算。在 k 次循环后, S_2 会获得最终的加密结果,并将其发送到 S_1 进行解密。

3 系统详细设计

3.1 图像加密

使用矩阵 $I_{n,n}$ 表示像素为 $n \times n$ 的原始图像 I, 矩阵的元素是一个 8 位的整数。数据所有者从 0 到 255 中随机选择 $n \times n$ 个整数以生成随机矩阵 $I_2(x, y)$,然后通过计算 $I_1(x,y)$ 来对 $I(x,y)$ 进行加密: $I_1(x,y) = I(x,y) + I_2(x,y)$ 。密文 I_1 和 I_2 分别被发送到 S_1 和 S_2。

3.2 关键点定位

服务器 S_i 接收到密文 I_i 后,通过对 I_i 进行卷积和下采样操作创建高斯空间 $L_i(x,y,\sigma)$, 即

$$L_1(x,y,\sigma) = G(x,y,\sigma) \otimes I_1(x,y). \tag{2}$$

其中, $G(x,y,\sigma)$ 是高斯函数, $G(x,y,\sigma) = (1/2\pi\sigma^2) e^{-(x^2+y^2)/2\sigma^2}$, $\otimes$ 是指卷积操作。对于原始图像,重复使用公式(2)进行卷积操作,并进行 2 倍的下采样。接下来, S_1 计算高斯差空间,即 $D_1(x,y,\sigma) = L_1(x,y,r\sigma) - L_1(x,y,\sigma)$。其中, r 是一个预先定义好的常数。与此同时, S_2 也会计算其高斯差空间 $D_2(x,y,\sigma)$。

在极值检测过程中,每一个样本点将会与其 26 个邻居进行比较,也会与高斯差空间 $D(x,y,\sigma)$ 中的相邻尺度进行比较。如果样本点均大于或者小于

邻居和相邻尺度,则该样本点会被选择作为候选关键点。为了判断 $D(x,y,\sigma)$ 和 $D(x,y+1,\sigma)$ 的大小,云服务器 S_1 和 S_2 分别计算 $\Delta D_1 = D_1(x,y,\sigma) - D_1(x,y+1,\sigma)$ 和 $\Delta D_2 = D_2(x,y,\sigma) - D_2(x,y+1,\sigma)$,然后采用 SCP 算法比较 ΔD_1 和 ΔD_2 的大小。当 $\Delta D_1 \geqslant \Delta D_2$,则认为 $D(x,y,\sigma) \geqslant D(x,y+1,\sigma)$;否则,$D(x,y,\sigma) < D(x,y+1,\sigma)$。

在确定了候选的关键点后,将进行边缘响应移除操作,以去除不稳定的关键点。对于云服务器 S_1 的关键点 $D_1(x,y,\sigma)$,其在 x 方向、y 方向和 xy 方向的偏导数如下所示:

$$R_{1xx} = D_1(x+1,y,\sigma) + D_1(x-1,y,\sigma) - 2D_1(x,y,\sigma);$$

$$R_{1yy} = D_1(x,y+1,\sigma) + D_1(x,y-1,\sigma) - 2D_1(x,y,\sigma);$$

$$R_{1xy} = 0.25(D_1(x+1,y+1,\sigma) - D_1(x+1,y-1,\sigma) - D_1(x-1,y+1,\sigma) + D_1(x-1,y-1,\sigma)).$$

同理可以计算云服务器 S_2 的关键点 $D_2(x,y,\sigma)$ 的三个方向上的偏导数 R_{2xx}、R_{2yy} 和 R_{2xy}。S_1 和 S_2 分别使用 SMP 协议计算 R_{1xx} 和 R_{2yy} 的乘积 $M_1(R_{1xx}R_{2yy})$ 和 $M_2(R_{1xx}R_{2yy})$,并有 $M_1(R_{1xx}R_{2yy}) + M_2(R_{1xx}R_{2yy}) = R_{1xx}R_{2yy}$。同理,$S_1$ 可以通过计算得到 $M_1(R_{1yy}R_{2xx})$ 和 $M_1(R_{1xy}R_{2xy})$,S_2 可以通过计算得到 $M_2(R_{1yy}R_{2xx})$ 和 $M_2(R_{1xy}R_{2xy})$。云服务器 S_1 中的海塞矩阵的迹和行列式如下所示:

$$Tr_1 = R_{1xx} + R_{1yy}$$

$$Det_1 = R_{1xx}R_{1yy} - R_{1xy}^2 - (M_1(R_{1xx}R_{2yy}) + M_1(R_{1yy}R_{2xx}) - 2M_1(R_{1xy}R_{2xy})).$$

同理可以计算服务器 S_2 中的海塞矩阵的迹 Tr_2 和行列式 Det_2。最后,为了判断点 $D_1(x,y,\sigma)$ 是否稳定,S_1 和 S_2 会将 Det_1 和 $-Det_2$ 进行比较:如果 $Det_1 \leqslant -Det_2$,则该点会被从候选关键点集剔除。若 $Det_1 > -Det_2$,则计算 $Z_1 = rTr_1^2 - 2rM_1(Tr_1Tr_2) - (r+1)^2Det_1$ 和 $Z_2 = -rTr_2^2 + 2rM_2(Tr_1Tr_2) + (r+1)^2Det_2$。如果 $Z_1 < Z_2$,则认为点 $D_1(x,y,\sigma)$ 是稳定的。

3.3　方向指定

一旦确定了关键点的位置,就应根据像素差异计算梯度幅度和方向,以便建立方向直方图。与参数 σ 相关联的关键点 $D(x,y,\sigma)$ 用于选择具有最接近尺度的高斯平滑图像 L,所有以下计算以尺度不变的方式执行。为了便于表达,将在以下讨论中省略参数 σ。

对于关键点 $L(x,y)$ 上的方向计算,S_1 和 S_2 将首先确定 $L(x,y+1)$ 与 $L(x,y-1)$ 之间,以及 $L(x+1,y)$ 和 $L(x-1,y)$ 之间的大小。当 $L(x,y+1) > L(x,y-1)$ 时,$\alpha = 1$;否则,$\alpha = -1$。当 $L(x+1,y) \geqslant L(x-1,y)$,$\beta = 1$;否则 $\beta = -1$。通过对 $L_1(x,y+1) - L_1(x,y-1)$ 和 $L_2(x,y+1) - L_2(x,y-1)$ 使用 SCP 协议,可以实现安全比较。为了获得更精细的方向范围,S_1 进行以下的计算:

$$\Delta Diff_1 = Diff_{1y} - kDiff_{1x} = \alpha(L_1(x,y+1) - L_1(x,y-1)) - k\beta(L_1(x+1,y) - L_1(x-1,y));$$

同理,S_2 也会计算 $\Delta Diff_2 = Diff_{2y} - kDiff_{2x}$,其中,常数 k 用来确定方向的间隔。然后 S_1 和 S_2 会使用 SCP 协议比较 $\Delta Diff_1$ 和 $\Delta Diff_2$ 的大小。如果 $\Delta Diff_1 \geqslant \Delta Diff_2$,则方向大于 $\arctan k$。重复上述步骤,可以得到一个粒度更小的方向度数范围。

将梯度的幅度定义为 $m(x,y) = |Diff_x| + |Diff_y|$,其中,$Diff_x = L(x+1,y) - L(x-1,y)$,$Diff_y = L(x,y+1) - L(x,y-1)$。$S_1$ 和 S_2 分别计算其梯度幅度 $m_1(x,y)$ 和 $m_2(x,y)$,然后在被相同的高斯窗口加权后,根据其方向分别累加 m_1 和 m_2,以构建加密方向直方图 H_1 和 H_2。

为了检测原始方向直方图 H 中的峰值,需要在两个区间之中找到累积梯度幅度更大的区间。假设 $\sum_{10^\circ} m_1$ 和 $\sum_{20^\circ} m_1$ 分别是 H_1 在方向范围为 0° ~ 10° 和 10° ~ 20° 的累积梯度幅度,$\sum_{10^\circ} m_2$ 和 $\sum_{20^\circ} m_2$ 分别是 H_2 在方向范围为 0° ~ 10° 和 10° ~ 20° 的累积梯度幅度。采用 SCP 协议比较 $\sum_{10^\circ} m_1 - \sum_{20^\circ} m_1$ 和 $\sum_{10^\circ} m_2 - \sum_{20^\circ} m_2$ 的大小,:如果前者较大,则认为 0° ~ 10° 的区间更大。重复上述步骤,就能找到全局和局部峰值。

3.4　图像描述符生成

S_1 和 S_2 检测到关键点上的主导方向后,将生成各自的描述符。首先,坐标和梯度方向相对于关键点方向旋转。在旋转过程中需要确定梯度方向的值,因此使用其方向所在区间的中值来替换实际的方向值。例如,如果一个点的方向位于 10° ~ 20° 的区间内,使用 15°作为其梯度方向值来执行旋转操作。这种近似的取向替代对最终结果的影响可以忽略不计[9]。其次,在关键点周围的 4×4 子区域中,采样点的梯度大小(即 $m(x,y)$)由高斯窗口加权并累积成 4×4 方向直方图。接下来,每个直方图由 8 个方向区间组成,每个区间的跨度为 45°。为了提

高系统效率,在方向分配阶段预先计算所有梯度。最后,采用具有 128 个维度的特征向量表示关键点的 16 个直方图。设 V_1 和 V_2 分别表示由 S_1 和 S_2 生成的特征向量,并会被发送到数据所有者 O。数据所有者通过计算 $V = V_1 - V_2$ 来恢复实际特征向量,其将被归一化为单位长度,以便实现仿射的变化的仿射变化的不变性。

4 实验评估

本节采用图像数据集 INRIA Graffiti 来评估安全 SIFT 外包方案的有效性和效率。实验环境配置如下:操作系统为 Linux Ubuntu,算法实现采用 C ++高级编程语言,处理器为 Intel 酷睿 3.1 GHz,内存为 8 GB。

将本文算法与 ED-SIFT 进行比较。为了更好地评估特征向量的独特性和鲁棒性,选择测量关键点匹配的性能,即给定一个点,将在数据集中找到该点的所有匹配。将计算特征向量之间的欧几里德距离,以找到数据集中的最近点和第二个最近的点。如果这两个点之间的距离比率低于阈值 t, 则该点与其最近点匹配。使用召回率和匹配准确度作为评估指标。图 1 是召回率和匹配准确度之间的关系。匹配准确度是准确匹配数量与总匹配数量的比。图 2 和图 3 分别是计算和通信开销的对比结果。结合三个实验结果图,可观察到本文的方法无论是在召回率、精确度及开销方面都要优于现有的方法。

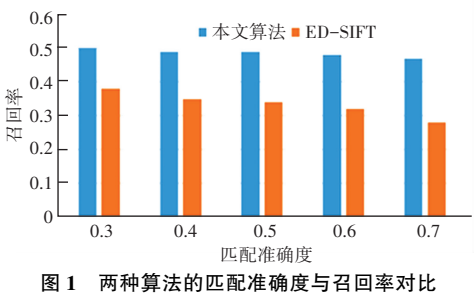

图 1 两种算法的匹配准确度与召回率对比

Fig. 1 The comparison between matching precision and recall of the two algorithms

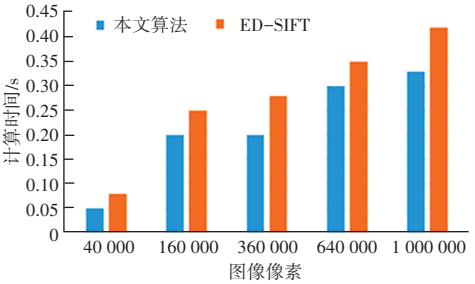

图 2 两种算法的计算开销对比

Fig. 2 The comparison of computing overhead of the two algorithms

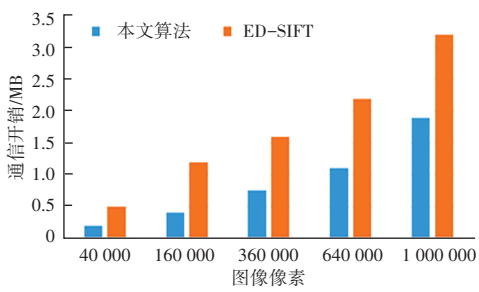

图 3 两种算法的通信开销对比

Fig. 3 The comparison of communication overhead of the two algorithms

5 结束语

本文提出了一种新颖的隐私保护图像特征提取方法,该方法由安全交互协议 SMP 和安全比较协议 SCP 组成。通过实验,分析评估了该方法的有效性,实验结果表明该方法优于现有技术。未来的工作集中于在真实的云计算集群中部署该方法,进一步评估该方法在真实云计算流量下的性能。

参考文献

[1] REN K, WANG C, WANG Q. Security challenges for the public cloud[J]. IEEE Internet Computing, 2012, 16(1): 69-73.

[2] XIA Z, WANG X, SUN X, et al. A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data[J]. IEEE transactions on parallel and distributed systems, 2015, 27(2): 340-352.

[3] HSU C Y, LU C S, PEI S C. Image feature extraction in encrypted domain with privacy-preserving SIFT[J]. IEEE transactions on image processing, 2012, 21(11): 4593-4607.

[4] QIN Z, YAN J, REN K, et al. Towards efficient privacy-preserving image feature extraction in cloud computing[C]// Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 497-506.

[5] WANG S, NASSAR M, ATALLAH M, et al. Secure and private outsourcing of shape-based feature extraction[C]//International conference on information and communications security. Springer, Cham, 2013: 90-99.

[6] HUANG Y, EVANS D, KATZ J, et al. Faster secure two-party computation using garbled circuits[C]//USENIX Security Symposium. 2011, 201(1): 331-335.

[7] WANG Q, HU S, REN K, et al. Catch me in the dark: Effective privacy-preserving outsourcing of feature extractions over image data[C]//IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE, 2016: 1-9.

[8] FAROKHI F, SHAMES I, BATTERHAM N. Secure and private control using semi-homomorphic encryption[J]. Control Engineering Practice, 2017, 67: 13-20.

[9] HU S, WANG Q, WANG J, et al. Securing SIFT: Privacy-preserving outsourcing computation of feature extractions over encrypted image data[J]. IEEE Transactions on Image Processing, 2016, 25(7): 3411-3425.