

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 4 月 16 日 (16.04.2020)



(10) 国际公布号
WO 2020/073519 A1

(51) 国际专利分类号:

H04L 29/06 (2006.01) *G10L 17/16* (2013.01)
G10L 17/00 (2013.01) *G10L 25/18* (2013.01)
G10L 17/02 (2013.01) *G10L 25/24* (2013.01)

(21) 国际申请号: PCT/CN2018/124402

(22) 国际申请日: 2019 年 2 月 1 日 (01.02.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

201811184775.3 2018 年 10 月 11 日 (11.10.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 杨翹楚(YANG, Qiaochu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市明日今典知识产权代理事务所(普通合伙)(SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL));
中国广东省深圳市南山区南头街道智恒新兴产业园 E 区 01B 栋 405 室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB,

(54) Title: VOICEPRINT VERIFICATION METHOD AND APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 声纹验证的方法、装置、计算机设备以及存储介质

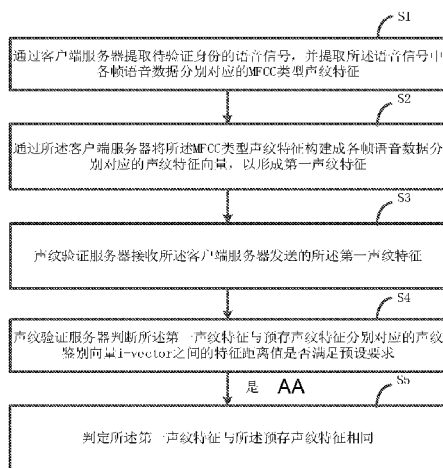


图 1

- S1 A client server extracts a voice signal to be subjected to identity verification, and extracts MFCC type voiceprint features respectively corresponding to frames of voice data in the voice signal
- S2 The client server constructs the MFCC type voiceprint features as voiceprint feature vectors respectively corresponding to the frames of voice data to form a first voiceprint feature
- S3 A voiceprint verification server receives the first voiceprint feature sent by the client server
- S4 The voiceprint verification server determines whether a feature distance value between voiceprint identification vectors (i-vectors) respectively corresponding to the first voiceprint feature and a pre-stored voiceprint feature meets a pre-set requirement
- S5 Determine that the first voiceprint feature is the same as the pre-stored voiceprint feature
- AA Yes

(57) Abstract: Disclosed is a voiceprint verification method. The method comprises: a client server extracting a voice signal to be subjected to identity verification, and extracting corresponding MFCC type voiceprint features, and constructing same as a first voiceprint feature corresponding to frames of voice data; a voiceprint verification server receiving the first voiceprint feature; the voiceprint verification server determining whether a feature distance value between voiceprint identification vectors (i-vectors) respectively corresponding to the first voiceprint feature and a pre-stored voiceprint feature meets a pre-set requirement; and if so, determining that the first voiceprint feature is the same as the pre-stored voiceprint feature.

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种声纹验证的方法, 包括: 通过客户端服务器提取待验证身份的语音信号, 并提取对应的MFCC类型声纹特征并构建成各帧语音数据对应的第一声纹特征; 声纹验证服务器接收第一声纹特征; 声纹验证服务器判断第一声纹特征与预存声纹特征分别对应的声纹鉴别向量*i-vector*之间的特征距离值是否满足预设要求; 若满足, 则判定第一声纹特征与预存声纹特征相同。

声纹验证的方法、装置、计算机设备以及存储介质

本申请要求于2018年10月11日提交中国专利局、申请号为2018111847753，发明名称为“声纹验证的方法、装置、计算机设备以及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请涉及到声纹验证领域，特别是涉及到声纹验证的方法、装置、计算机设备以及存储介质。

背景技术

目前，很多大型金融公司的业务范围涉及保险、银行、投资等多个业务范畴，而每个业务范畴通常都需要同客户进行沟通，且都需要进行反欺诈识别，因此，对客户的身​​份验证及反欺诈识别也就成为保证业务安全的重要组成部分。在客户身份验证环节中，声纹验证由于其具有的实时性和方便性而被许多公司采用。客户声纹模型的训练和客户身份的验证需要采集客户的语音数据，而语音数据的获得往往来源于金融公司与客户的谈话录音。发明人意识到，由于商业洽谈往往涉及机密内容，将语音数据由网络传输到后台再进行语音特征参数的提取不利于数据保密性。

技术问题

本申请的主要目的为提供声纹验证的方法，旨在解决现有声纹验证过程中需将客户端采集的语音数据发送至后台进行声纹特征提取，导致传输中语音数据的保密性较差的技术问题。

技术解决方案

本申请提出一种声纹验证的方法，包括：

通过客户端服务器提取待验证身份的语音信号，并提取所述语音信号中各帧语音数据分别对应的 MFCC 类型声纹特征；

通过所述客户端服务器将所述 MFCC 类型声纹特征构建成各帧语音数据分别对应的声纹特征向量，以形成第一声纹特征；

声纹验证服务器接收所述客户端服务器发送的所述第一声纹特征；

声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求；

若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

本申请还提供了一种声纹验证系统，包括客户端、客户端服务器和声纹验

证服务器；

所述客户端采集待验证身份的语音信号，并将所述语音信号发送到所述客户端服务器；

所述客户端服务器接收所述语音信号，并对所述语音信号进行声纹特征提取得到第一声纹特征，将第一声纹特征传输至声纹验证服务器；

所述声纹验证服务器接收所述第一声纹特征，并将所述第一声纹特征与预存声纹特征进行比较分析，以判断所述第一声纹特征与所述预存声纹特征是否相同，并将判断结果反馈至所述客户端服务器；

所述客户端服务器根据所述判断结果控制所述客户端进行反馈响应。

本申请还提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机程序，所述处理器执行所述计算机程序时实现上述方法的步骤。
本申请还提供了一种计算机非易失性可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述的方法的步骤。

有益效果

本申请将声纹特征向量提取的功能前置到客户端服务器上完成，客户端通过录音采集语音信号后直接在本地的客户端服务器提取语音信号的声纹特征向量，然后再将声纹特征向量传输至第三方技术支持的验证服务器上进行声纹验证，声纹验证模型的训练和说话人辨认过程，由于声纹特征向量无法再反推还原为语音信号的原始数据，有利于对客户录音的语音信号进行数据保密，提高数据安全性，使客户身份认证流程的安全性得到了提高。本申请通过提取声纹特征向量后的数据传输至服务器进行声纹验证，声纹特征向量数据比原始语音信号数据更为轻便，大大增加了传输效率。本申请基于GMM-UBM实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量i-vector，降低计算成本，降低声纹验证的使用成本。在验证过程中通过与多人的预存数据进行比较分析，降低声纹验证的等错率，降低声纹验证的模型误差带来的影响。

附图说明

图1 本申请一实施例的声纹验证的方法流程示意图；

图2 本申请一实施例的计算机设备内部结构示意图。

本发明的最佳实施方式

参照图1，本申请一实施例的声纹验证的方法，通过客户端采集信息，通过服务器进行声纹验证，方法包括：

S1: 通过客户端服务器提取待验证身份的语音信号, 并提取所述语音信号中各帧语音数据分别对应的MFCC类型声纹特征。

本实施例的MFCC (Mel Frequency Cepstrum Coefficient, 梅尔频率倒谱系数) 类型声纹特征具有非线性特征, 使客户的语音信号在各频段上的分析结果更贴近人体发出的真实语音的特征, 提高声纹验证的效果。

S2: 通过所述客户端服务器将所述MFCC类型声纹特征构建成各帧语音数据分别对应的声纹特征向量, 以形成第一声纹特征。

本实施例根据提取的MFCC类型声纹特征构建各帧语音数据分别对应的声纹特征向量, 然后通过语音信号的各帧语音数据的排序, 将分别对应的MFCC类型声纹特征组合在一起, 得到客户的语音信号对应的第一声纹特征, 上述构建过程依然在客户端服务器完成, 以增强数据传输过程中的数据保密性。

S3: 声纹验证服务器接收所述客户端服务器发送的所述第一声纹特征。

本实施例将第一声纹特征的提取工作前置到客户端服务器完成, 以便客户端服务器接收录音采集的客户的语音信号后, 直接在客户端服务器提取语音信号对应的第一声纹特征, 然后再传输至第三方技术支持的声纹验证服务器进行声纹验证。由于第一声纹特征无法再通过反推还原为原始的语音信号, 有利于对客户录音的语音信号进行数据保密, 提高数据安全性, 使客户身份认证流程的安全性得到了提高, 同时, 第一声纹特征比语音信号数据量更小, 大大增加了传输效率。通过客户端服务器对采集的语音信号提取声纹特征, 将提取后的声纹特征传输至声纹验证服务器进行声纹验证, 使声纹特征提取的客户端服务器和声纹验证服务器进行分离。

S4: 声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量i-vector之间的特征距离值是否满足预设要求。

本实施例的预设要求包括特征距离值达到指定的预设阈值范围等, 可根据具体的应用场景进行自定义设定, 以更广泛地满足个性化使用需求。

S5: 若满足, 则判定第一声纹特征与预存声纹特征相同, 否则不相同。

本实施例将判定所述第一声纹特征与所述预存声纹特征相同, 则通过服务器向客户端反馈验证通过的结果到客户端, 否则, 反馈验证失败的结果到客户端, 以便客户端根据反馈结果进行进一步的应用操作。举例地, 验证通过后控制智能门打开等。再举例地, 验证失败指定次数后控制安全系统进行锁屏, 以防犯罪分子进一步破坏电子银行系统。

进一步地，本实施例的步骤S4，包括：

S41：将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量i-vector。

本实施例基于GMM-UBM (Gaussian Mixture Model-Universal Background Model, 高斯混合模型-背景模型) 实现将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量i-vector。本实施例的GMM-UBM的训练过程如下：B1：获取预设数量（例如，10万个）的语音数据样本，每个语音数据样本对应一个声纹鉴别向量，每个语音样本可以采集自不同的人在不同环境中的语音，这样的语音数据样本用来训练能够表征一般语音特性的通用背景模型（GMM-UBM）；B2、分别对各个语音数据样本进行处理以提取出各个语音数据样本对应的预设类型声纹特征，并基于各个语音数据样本对应的预设类型声纹特征构建各个语音数据样本对应的声纹特征向量；B3、将构建出的所有预设类型声纹特征向量分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比之后小于或等于100%；B4、利用训练集中的声纹特征向量对所述第二模型进行训练，并在训练完成之后利用验证集对训练的所述第二模型的准确率进行验证；B5、若准确率大于预设准确率（例如，98.5%），则模型训练结束，否则，增加语音数据样本的数量，并基于增加后的语音数据样本重新执行上述步骤B2、B3、B4、B5。

本实施例的声纹鉴别向量采用声纹鉴别向量i-vector表达，声纹鉴别向量i-vector是一个向量，相对于高斯空间的维度来讲，声纹鉴别向量i-vector维度更低，便于降低计算成本。

S42：通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x| |y|}$ ，计算第一声纹特征对应的声纹鉴别向量i-vector与预存声纹特征对应的声纹鉴别向量i-vector之间的距离值，其中，x代表预存声纹特征对应的声纹鉴别向量i-vector，y代表第一声纹特征对应的声纹鉴别向量i-vector。

S43：判断所述余弦距离值是否满足预设条件。

本实施例的预设条件包括余弦距离值在指定的阈值范围内等，可根据需要设定。本实施例通过将预存的多个人的声纹特征数据中各自对应的预存声纹特征与所述第一声纹特征分别计算的第一余弦距离值进行从小到大排序，判断预设排序在前的几个第一余弦距离值中是否包括目标人的预存声纹特征对应的

第一余弦距离值，若包括则判定余弦距离值满足预设条件。本申请另一实施例通过判断目标人的预存声纹特征与所述第一声纹特征之间的第二余弦距离值是否小于或等于预设阈值，若小于或等于，则判定余弦距离值满足预设条件。

S44：若所述余弦距离值满足预设条件，则判定所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量i-vector之间的特征距离值满足预设要求，
5 否则不满足预设要求。

进一步地，本实施例的步骤S41，包括：

S410：将提取得到的各帧语音数据分别对应的声纹特征向量分别输入到GMM-UBM模型，得到表征各帧语音数据在各高斯分量上的概率分布的高斯超向
10 量。

S411：将各所述高斯超向量利用公式 $m_r = \mu + T\omega_r$ ，计算得到各帧语音数据分别对应的低维度的声纹鉴别向量i-vector，其中 m_r 为各帧语音数据的高斯超向量， μ 为所述GMM-UBM模型的均值超向量，T为各帧语音数据的低维度的声纹鉴别向量i-vector， ω_r 为映射到高维度的高斯空间的转换矩阵。

本实施例的T训练采用EM算法。EM算法，指的是最大期望算法(Expectation Maximization Algorithm，又译期望最大化算法)，是一种迭代算法，在统计学中被用于寻找，依赖于不可观察的隐性变量的概率模型中，参数的最大似然估计。最大期望算法经过两个步骤交替进行计算：1) 计算期望(E)，利用概率模型参数的现有估计值，计算隐藏变量的期望；2) 最大化(M)，利用E步
15 上求得的隐藏变量的期望，对参数模型进行最大似然估计。上步找到的参数估计值被用于下步计算中，不断交替进行。

进一步地，本实施例的步骤S43，包括：

S430：分别获取预存的多个人的声纹特征数据中各自对应的预存声纹特征与
25 所述第一声纹特征之间的第一余弦距离值，其中，多个人的声纹特征数据中包括目标人的预存声纹特征。

本实施例通过将预存的包括目标人的多人的声纹特征数据，同时用于判断当前采集的语音信号的声纹特征是否与目标人的声纹特征相同，以提高判断准

确性。本实施例通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x||y|}$ ，表示各所述预存声纹特

征与所述第一声纹特征之间的第一余弦距离值，其中，x代表各预存声纹鉴别
30 向量，y代表第一声纹特征的声纹鉴别向量i-vector，余弦距离值越小，表明

两声纹特征更接近或相同。本实施例的“第一”，仅用作区别，不用于限定，其他处的作用相同，不赘述。

S431: 将各所述第一余弦距离值按照从小到大的顺序进行排序。

本实施例通过将各所述预存声纹特征与所述第一声纹特征之间的第一余弦距离值进行从小到大排序，以便更准确地分析第一声纹特征与各预存声纹特征的相似度分布状态，以便更准确地获得对第一声纹特征的验证。

S432: 判断排序在前的预设数量的第一余弦距离值中，是否包括所述目标人的预存声纹特征对应的第一余弦距离值。

本实施例通过排序在前的预设数量的第一余弦距离值中包括所述目标人的预存声纹特征对应的第一余弦距离值，则判定第一声纹特征与预存的目标人的声纹特征相同，以减小模型误差带来的识别等错率，上述等错率为“应验证通过时发生的验证未通过的频率，与应验证未通过时发生的验证通过的频率相等”。本实施例的预设数量的第一余弦距离值包括1个、2个或3个等，可根据使用需求进行自设定。

S433: 若是，则判定余弦距离值满足预设条件，否则不满足预设条件。

进一步地，本申请另一实施例的步骤S43，包括：

S434: 获取目标人的预存声纹特征与第一声纹特征间的第二余弦距离值。

本实施例通过只针对性地比较一个第二余弦距离值，减小比较计算量，提高验证速率。

S435: 判断所述第二余弦距离值是否小于或等于预设阈值。

本实施例通过设定第一声纹特征与目标用户的预存声纹特征的距离阈值，实现有效的声纹验证。举例地，预设阈值为0.6。

S436: 若是，则判定余弦距离值满足预设条件，否则不满足预设条件。

本实施例计算第一声纹特征与目标用户的预存声纹特征的余弦距离小于或等于预设阈值，则判定余弦距离值满足预设条件，确定第一声纹特征与目标用户的预存声纹特征相同，则验证通过；若计算第一声纹特征与目标用户的预存声纹特征的余弦距离大于预设阈值，则判定所述距离值不满足预设条件，确定第一声纹特征与目标用户的预存声纹特征不相同，则验证失败。

本申请还提供了一种声纹验证系统，包括客户端、客户端服务器和声纹验证服务器；

客户端采集待验证身份的语音信号，并将语音信号发送到客户端服务器；

所述客户端服务器接收所述语音信号,并对所述语音信号进行声纹特征提取得到第一声纹特征,将第一声纹特征传输至声纹验证服务器;

所述声纹验证服务器接收所述第一声纹特征,并将所述第一声纹特征与预存声纹特征进行比较分析,以判断所述第一声纹特征与所述预存声纹特征是否相同,并将判断结果反馈至所述客户端服务器;

所述客户端服务器根据所述判断结果控制所述客户端进行反馈响应。

进一步地,本实施例的所述语音信号的连续模拟信号通过客户端按照指定采样周期进行采样,以形成离散模拟信号,并指定编码规则量化为数字信号;所述客户端服务器接收所述语音信号,并对所述语音信号进行声纹特征提取得到第一声纹特征的过程如下:

S101,所述客户端服务器将所述数字信号进行预加重后,对预加重的数字信号进行分帧处理,得到各帧语音数据;S102,根据
$$f_{Mel} = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right)$$
将各帧语音数据从线性频谱域映射到梅尔频谱域,其中, f_{Mel} 表示梅尔频谱值, f 表示线性频谱值;S103,将转化为梅尔频谱域的各帧语音数据输入到一组梅尔三角滤波器组,计算每个频段的梅尔三角滤波器输出的对数能量,得到各帧语音数据分别对应的对数能量序列;S104,将各所述对数能量序列进行离散余弦变换,得到各帧语音数据分别对应的MFCC类型声纹特征;将所述MFCC类型声纹特征构建成各帧语音数据分别对应的声纹特征向量,以形成所述第一声纹特征。

上述预加重,由于人体的生理特性,语音信号的高频成分往往被压抑,预加重的作用是补偿高频成分;上述分帧处理中,由于语音信号的“瞬时平稳性”,在进行频谱分析时对一段语音信号进行分帧处理(一般为10至30毫秒一帧),然后以帧为单位进行特征提取;上述分帧处理后进行了加窗处理,作用是减少帧起始和结束地方信号的不连续性问题,本实施例采用汉明窗进行加窗处理。

本实施例将声纹特征向量提取的功能前置到客户端服务器上完成,客户端通过录音采集语音信号后直接在本地的客户端服务器提取语音信号的声纹特征向量,然后再将声纹特征向量传输至第三方技术支持的验证服务器上进行声纹验证,声纹验证模型的训练和说话人辨认过程,由于声纹特征向量无法再反推还原为语音信号的原始数据,有利于对客户录音的语音信号进行数据保密,提高数据安全性,使客户身份认证流程的安全性得到了提高。本实施例通过提

取声纹特征向量后的数据传输至服务器进行声纹验证,声纹特征向量数据比原始语音信号数据更为轻便,大大增加了传输效率。本实施例基于GMM-UBM实现将各所述声纹特征向量分别映射为低维度的声纹鉴别向量i-vector,降低计算成本,降低声纹验证的使用成本。在验证过程中通过与多人的预存数据进行比较分析,降低声纹验证的等错率,降低声纹验证的模型误差带来的影响。

进一步地,判断结果包括第一声纹特征与预存声纹特征不相同,所述客户端服务器根据所述判断结果控制所述客户端进行反馈响应的过程,包括:

客户端服务器生成身份验证不成功的反馈信息并发送至所述客户端;

判断预设时间内根据所述第一声纹特征生成身份验证不成功的反馈信息的次数,是否超过预设次数。

若超过预设次数,则控制所述客户端处于禁用状态,并发出警报。

本声纹验证系统包括警报和安全管控装置,以增强该声纹验证系统在实际应用过程的功能完备性,提高管理安全和信息安全。

参照图2,本申请实施例中还提供一种计算机设备,该计算机设备可以是服务器,其内部结构可以如图2所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储声纹验证等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令在执行时,执行如上述各方法的实施例的流程。本领域技术人员可以理解,图2中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定。

本申请一实施例还提供一种计算机非易失性可读存储介质,其上存储有计算机可读指令,该计算机可读指令在执行时,执行如上述各方法的实施例的流程。以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本申请的专利保护范围内。

权利要求书

1、一种声纹验证的方法，其特征在于，包括：

通过客户端服务器提取待验证身份的语音信号，并提取所述语音信号中各帧语音数据分别对应的 MFCC 类型声纹特征；

5 通过所述客户端服务器将所述 MFCC 类型声纹特征构建成各帧语音数据分别对应的声纹特征向量，以形成第一声纹特征；

声纹验证服务器接收所述客户端服务器发送的所述第一声纹特征；

声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求；

10 若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

2、根据权利要求 1 所述的声纹验证的方法，其特征在于，所述声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求的步骤，包括：

15 将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量 i-vector；

通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x| |y|}$ ，计算第一声纹特征对应的声纹鉴别

向量 i-vector 与预存声纹特征对应的声纹鉴别向量 i-vector 之间的余弦距离值 $\cos(x, y)$ ，其中，x 代表预存声纹特征对应的声纹鉴别向量 i-vector，y 代表第一声纹特征对应的声纹鉴别向量 i-vector；

20 判断所述余弦距离值是否满足预设条件；

若满足，则判定所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值满足预设要求，否则不满足预设要求。

3、根据权利要求 2 所述的声纹验证的方法，其特征在于，所述将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量 i-vector 的步骤，包括：

25 将提取得到的各帧语音数据分别对应的声纹特征向量分别输入到 GMM-UBM 模型，得到表征各帧语音数据在各高斯分量上的概率分布的高斯超向量；

30 将各所述高斯超向量利用公式 $m_r = \mu + T\omega_r$ ，计算得到各帧语音数据分别对应的低维度的声纹鉴别向量 i-vector，其中 m_r 为各帧语音数据的高斯超

向量, μ 为所述 GMM-UBM 模型的均值超向量, T 为各帧语音数据的低维度的声纹鉴别向量 i-vector, ω_r 为映射到高维度的高斯空间的转换矩阵。

4、根据权利要求 2 所述的声纹验证的方法, 其特征在于, 所述判断所述余弦距离值是否满足预设条件的步骤, 包括:

5 分别获取预存的多个人的声纹特征数据中各自对应的预存声纹特征与所述第一声纹特征之间的第一余弦距离值, 其中, 多个人的声纹特征数据中包括目标人的预存声纹特征;

将各所述第一余弦距离值按照从小到大的顺序进行排序;

10 判断排序在前的预设数量的第一余弦距离值中, 是否包括所述目标人的预存声纹特征对应的第一余弦距离值;

若是, 则判定所述第一余弦距离值满足预设条件, 否则不满足预设条件。

5、根据权利要求 2 所述的声纹验证的方法, 其特征在于, 所述判断所述余弦距离值是否满足预设条件的步骤, 包括:

15 获取目标人的预存声纹特征与所述第一声纹特征之间的第二余弦距离值; 判断所述第二余弦距离值是否小于或等于预设阈值;

若是, 则判定所述第二余弦距离值满足预设条件, 否则不满足预设条件。

6、一种声纹验证系统, 其特征在于, 包括客户端、客户端服务器和声纹验证服务器;

20 所述客户端采集待验证身份的语音信号, 并将所述语音信号发送到所述客户端服务器;

所述客户端服务器接收所述语音信号, 并对所述语音信号进行声纹特征提取得到第一声纹特征, 将第一声纹特征传输至声纹验证服务器;

25 所述声纹验证服务器接收所述第一声纹特征, 并将所述第一声纹特征与预存声纹特征进行比较分析, 以判断所述第一声纹特征与所述预存声纹特征是否相同, 并将判断结果反馈至所述客户端服务器;

所述客户端服务器根据所述判断结果控制所述客户端进行反馈响应。

7、根据权利要求 6 所述的声纹验证系统, 其特征在于, 所述语音信号的连续模拟信号通过客户端按照指定采样周期进行采样, 以形成离散模拟信号, 30 并通过指定编码规则量化为数字信号; 所述客户端服务器接收所述语音信号, 并对所述语音信号进行声纹特征提取得到第一声纹特征的过程, 包括:

所述客户端服务器将所述数字信号进行预加重后,对预加重的数字信号进行分帧处理,得到各帧语音数据;

根据 $f_{Mel} = 2595 \times \log_{10}(1 + \frac{f}{700})$ 将所述各帧语音数据从线性频谱域映射到梅尔频谱域,其中, f_{Mel} 表示梅尔频谱值, f 表示线性频谱值;

5 将转化为梅尔频谱域的各帧语音数据输入到梅尔三角滤波器组,计算每个频段的梅尔三角滤波器输出的对数能量,得到各帧语音数据分别对应的对数能量序列;

将各所述对数能量序列进行离散余弦变换,得到各帧语音数据分别对应的 MFCC 类型声纹特征;

10 将所述 MFCC 类型声纹特征构建成各帧语音数据分别对应的声纹特征向量,以形成所述第一声纹特征。

8、根据权利要求 6 所述的声纹验证系统,其特征在于,所述判断结果包括所述第一声纹特征与所述预存声纹特征不相同,所述客户端服务器根据所述判断结果控制所述客户端进行反馈响应的过程,包括:

15 客户端服务器生成身份验证不成功的反馈信息并发送至所述客户端;

判断预设时间内根据所述第一声纹特征生成身份验证不成功的反馈信息的次数,是否超过预设次数。

若超过预设次数,则控制所述客户端处于禁用状态,并发出警报。

20 9、一种计算机设备,包括存储器和处理器,所述存储器存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现声纹验证的方法,声纹验证的方法包括:

通过客户端服务器提取待验证身份的语音信号,并提取所述语音信号中各帧语音数据分别对应的 MFCC 类型声纹特征;

25 通过所述客户端服务器将所述 MFCC 类型声纹特征构建成各帧语音数据分别对应的声纹特征向量,以形成第一声纹特征;

声纹验证服务器接收所述客户端服务器发送的所述第一声纹特征;

声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求;

若满足,则判定所述第一声纹特征与所述预存声纹特征相同,否则不相同。

30 10、根据权利要求 9 所述的计算机设备,其特征在于,所述声纹验证服务

器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求的步骤, 包括:

将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量 i-vector;

5 通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{|x| |y|}$, 计算第一声纹特征对应的声纹鉴别

向量 i-vector 与预存声纹特征对应的声纹鉴别向量 i-vector 之间的余弦距离值 $\cos(x, y)$, 其中, x 代表预存声纹特征对应的声纹鉴别向量 i-vector, y 代表第一声纹特征对应的声纹鉴别向量 i-vector;

判断所述余弦距离值是否满足预设条件;

10 若满足, 则判定所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值满足预设要求, 否则不满足预设要求。

11、根据权利要求 10 所述的计算机设备, 其特征在于, 所述将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量 i-vector 的步骤, 包括:

15 将提取得到的各帧语音数据分别对应的声纹特征向量分别输入到 GMM-UBM 模型, 得到表征各帧语音数据在各高斯分量上的概率分布的高斯超向量;

将各所述高斯超向量利用公式 $m_r = \mu + T\omega_r$, 计算得到各帧语音数据分别对应的低维度的声纹鉴别向量 i-vector, 其中 m_r 为各帧语音数据的高斯超
20 向量, μ 为所述 GMM-UBM 模型的均值超向量, T 为各帧语音数据的低维度的声纹鉴别向量 i-vector, ω_r 为映射到高维度的高斯空间的转换矩阵。

12、根据权利要求 10 所述的计算机设备, 其特征在于, 所述判断所述余弦距离值是否满足预设条件的步骤, 包括:

25 分别获取预存的多个人的声纹特征数据中各自对应的预存声纹特征与所述第一声纹特征之间的第一余弦距离值, 其中, 多个人的声纹特征数据中包括目标人的预存声纹特征;

将各所述第一余弦距离值按照从小到大的顺序进行排序;

判断排序在前的预设数量的第一余弦距离值中, 是否包括所述目标人的预存声纹特征对应的第一余弦距离值;

30 若是, 则判定所述第一余弦距离值满足预设条件, 否则不满足预设条件。

13、根据权利要求 10 所述的计算机设备，其特征在于，所述判断所述余弦距离值是否满足预设条件的步骤，包括：

获取目标人的预存声纹特征与所述第一声纹特征之间的第二余弦距离值；
判断所述第二余弦距离值是否小于或等于预设阈值；

5 若是，则判定所述第二余弦距离值满足预设条件，否则不满足预设条件。

14、一种计算机非易失性可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现声纹验证的方法，声纹验证的方法包括：

通过客户端服务器提取待验证身份的语音信号，并提取所述语音信号中各
10 帧语音数据分别对应的 MFCC 类型声纹特征；

通过所述客户端服务器将所述 MFCC 类型声纹特征构建成各帧语音数据分别对应的声纹特征向量，以形成第一声纹特征；

声纹验证服务器接收所述客户端服务器发送的所述第一声纹特征；

15 声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求；

若满足，则判定所述第一声纹特征与所述预存声纹特征相同，否则不相同。

15、根据权利要求 14 所述的计算机非易失性可读存储介质，其特征在于，所述声纹验证服务器判断所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值是否满足预设要求的步骤，包括：

20 将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别向量 i-vector；

通过余弦距离公式 $\cos(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$ ，计算第一声纹特征对应的声纹鉴别

向量 i-vector 与预存声纹特征对应的声纹鉴别向量 i-vector 之间的余弦距离值 $\cos(x, y)$ ，其中，x 代表预存声纹特征对应的声纹鉴别向量 i-vector，y
25 代表第一声纹特征对应的声纹鉴别向量 i-vector；

判断所述余弦距离值是否满足预设条件；

若满足，则判定所述第一声纹特征与预存声纹特征分别对应的声纹鉴别向量 i-vector 之间的特征距离值满足预设要求，否则不满足预设要求。

16、根据权利要求 15 所述的计算机非易失性可读存储介质，其特征在于，
30 所述将各帧语音数据分别对应的声纹特征向量分别映射为低维度的声纹鉴别

向量 i-vector 的步骤, 包括:

将提取得到的各帧语音数据分别对应的声纹特征向量分别输入到 GMM-UBM 模型, 得到表征各帧语音数据在各高斯分量上的概率分布的高斯超向量;

- 5 将各所述高斯超向量利用公式 $m_r = \mu + T\omega_r$, 计算得到各帧语音数据分别对应的低维度的声纹鉴别向量 i-vector, 其中 m_r 为各帧语音数据的高斯超向量, μ 为所述 GMM-UBM 模型的均值超向量, T 为各帧语音数据的低维度的声纹鉴别向量 i-vector, ω_r 为映射到高维度的高斯空间的转换矩阵。

- 10 17、根据权利要求 15 所述的计算机非易失性可读存储介质, 其特征在于, 所述判断所述余弦距离值是否满足预设条件的步骤, 包括:

分别获取预存的多个人的声纹特征数据中各自对应的预存声纹特征与所述第一声纹特征之间的第一余弦距离值, 其中, 多个人的声纹特征数据中包括目标人的预存声纹特征;

将各所述第一余弦距离值按照从小到大的顺序进行排序;

- 15 判断排序在前的预设数量的第一余弦距离值中, 是否包括所述目标人的预存声纹特征对应的第一余弦距离值;

若是, 则判定所述第一余弦距离值满足预设条件, 否则不满足预设条件。

- 18、根据权利要求 15 所述的计算机非易失性可读存储介质, 其特征在于, 所述判断所述余弦距离值是否满足预设条件的步骤, 包括:

- 20 获取目标人的预存声纹特征与所述第一声纹特征之间的第二余弦距离值;
判断所述第二余弦距离值是否小于或等于预设阈值;

若是, 则判定所述第二余弦距离值满足预设条件, 否则不满足预设条件。

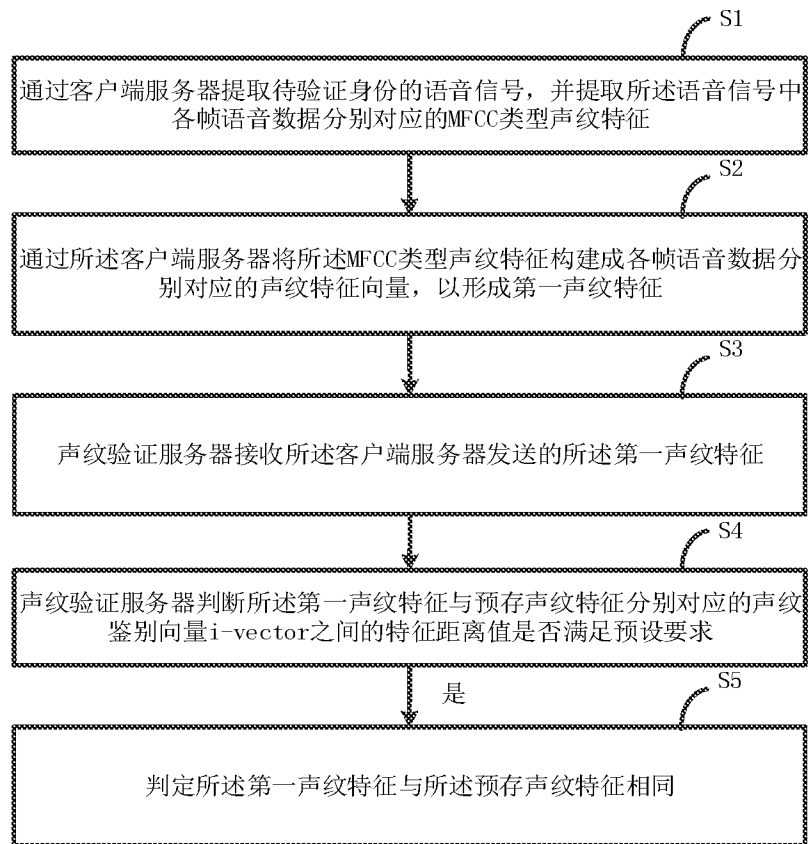


图 1

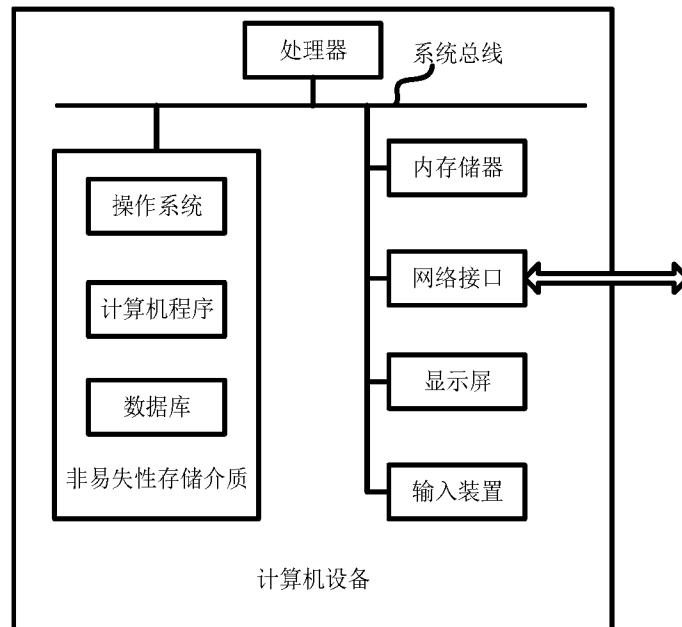


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/124402

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06(2006.01)i; G10L 17/00(2013.01)i; G10L 17/02(2013.01)i; G10L 17/16(2013.01)i; G10L 25/18(2013.01)i; G10L 25/24(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04L G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; CNKI; VEN; USTXT; EPTXT; WOTXT: 声纹, 身份, 验证, 认证, 语音, 提取, 获得, 特征向量, 服务器, 客户端, 匹配, 判断, 余弦, 距离, MFCC, feature, vector, GMM, UBM, voiceprint, authentication, verification, match, distance, server, client, determine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109257362 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 22 January 2019 (2019-01-22) description, paragraphs [0048]-[0115]	1-18
Y	CN 107993071 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 04 May 2018 (2018-05-04) description, paragraphs [0047]-[0145]	1-18
Y	CN 104680375 A (OPTIMIZED TECHNOLOGY SUZHOU CO., LTD.) 03 June 2015 (2015-06-03) description, paragraphs [0006]-[0030], and figures 2 and 3	1-18
A	CN 103955677 A (NANJING UNIVERSITY) 30 July 2014 (2014-07-30) entire document	1-18
A	US 2018158464 A1 (VERINT SYSTEMS LTD) 07 June 2018 (2018-06-07) entire document	1-18

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

31 May 2019

Date of mailing of the international search report

18 June 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration; PRC (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/124402

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109257362	A	22 January 2019	None			
CN	107993071	A	04 May 2018	None			
CN	104680375	A	03 June 2015	None			
CN	103955677	A	30 July 2014	CN	103955677	B	24 May 2017
US	2018158464	A1	07 June 2018	US	10109280	B2	23 October 2018

国际检索报告

国际申请号

PCT/CN2018/124402

A. 主题的分类 H04L 29/06(2006.01)i; G10L 17/00(2013.01)i; G10L 17/02(2013.01)i; G10L 17/16(2013.01)i; G10L 25/18(2013.01)i; G10L 25/24(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																				
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04L G10L 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS;CNTXT;CNKI;VEN;USTXT;EPTXT;WOTXT: 声纹, 身份, 验证, 认证, 语音, 提取, 获得, 特征向量, 服务器, 客户端, 匹配, 判断, 余弦, 距离, MFCC, feature, vector, GMM, UBM, voiceprint, authentication, verification, match, distance, server, client, determine																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109257362 A (平安科技深圳有限公司) 2019年 1月 22日 (2019 - 01 - 22) 说明书第[0048]-[0115]段</td> <td>1-18</td> </tr> <tr> <td>Y</td> <td>CN 107993071 A (平安科技深圳有限公司) 2018年 5月 4日 (2018 - 05 - 04) 说明书第[0047]-[0145]段</td> <td>1-18</td> </tr> <tr> <td>Y</td> <td>CN 104680375 A (优化科技苏州有限公司) 2015年 6月 3日 (2015 - 06 - 03) 说明书第[0006]-[0030]段, 图2、图3</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>CN 103955677 A (南京大学) 2014年 7月 30日 (2014 - 07 - 30) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>US 2018158464 A1 (VERINT SYSTEMS LTD) 2018年 6月 7日 (2018 - 06 - 07) 全文</td> <td>1-18</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109257362 A (平安科技深圳有限公司) 2019年 1月 22日 (2019 - 01 - 22) 说明书第[0048]-[0115]段	1-18	Y	CN 107993071 A (平安科技深圳有限公司) 2018年 5月 4日 (2018 - 05 - 04) 说明书第[0047]-[0145]段	1-18	Y	CN 104680375 A (优化科技苏州有限公司) 2015年 6月 3日 (2015 - 06 - 03) 说明书第[0006]-[0030]段, 图2、图3	1-18	A	CN 103955677 A (南京大学) 2014年 7月 30日 (2014 - 07 - 30) 全文	1-18	A	US 2018158464 A1 (VERINT SYSTEMS LTD) 2018年 6月 7日 (2018 - 06 - 07) 全文	1-18
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 109257362 A (平安科技深圳有限公司) 2019年 1月 22日 (2019 - 01 - 22) 说明书第[0048]-[0115]段	1-18																		
Y	CN 107993071 A (平安科技深圳有限公司) 2018年 5月 4日 (2018 - 05 - 04) 说明书第[0047]-[0145]段	1-18																		
Y	CN 104680375 A (优化科技苏州有限公司) 2015年 6月 3日 (2015 - 06 - 03) 说明书第[0006]-[0030]段, 图2、图3	1-18																		
A	CN 103955677 A (南京大学) 2014年 7月 30日 (2014 - 07 - 30) 全文	1-18																		
A	US 2018158464 A1 (VERINT SYSTEMS LTD) 2018年 6月 7日 (2018 - 06 - 07) 全文	1-18																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																				
国际检索实际完成的日期 2019年 5月 31日		国际检索报告邮寄日期 2020年 3月 5日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		受权官员 孙慧珠 电话号码 86-(512)-88996093																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/124402

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	109257362	A	2019年 1月 22日	无	
CN	107993071	A	2018年 5月 4日	无	
CN	104680375	A	2015年 6月 3日	无	
CN	103955677	A	2014年 7月 30日	CN 103955677	B 2017年 5月 24日
US	2018158464	A1	2018年 6月 7日	US 10109280	B2 2018年 10月 23日