



(12) 发明专利

(10) 授权公告号 CN 113408557 B

(45) 授权公告日 2023. 10. 13

(21) 申请号 202010184910.5

G06Q 10/105 (2023.01)

(22) 申请日 2020.03.17

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 113408557 A

CN 108563675 A, 2018.09.21

CN 109241378 A, 2019.01.18

CN 109710780 A, 2019.05.03

(43) 申请公布日 2021.09.17

CN 109800668 A, 2019.05.24

(73) 专利权人 深圳云天励飞技术有限公司
地址 518000 广东省深圳市龙岗区园山街
道龙岗大道8288号深圳大运软件小镇
17栋1楼

US 2016267173 A1, 2016.09.15

CN 109800674 A, 2019.05.24

CN 109815829 A, 2019.05.28

CN 110765134 A, 2020.02.07

CN 110334232 A, 2019.10.15

US 2017109588 A1, 2017.04.20

(72) 发明人 戴世稳

审查员 戴琦琦

(74) 专利代理机构 深圳驿航知识产权代理事务
所(普通合伙) 44605

专利代理师 杨伦

(51) Int. Cl.

G06V 10/74 (2022.01)

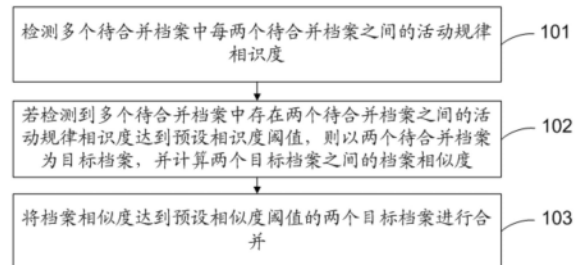
权利要求书2页 说明书12页 附图4页

(54) 发明名称

一种档案合并方法、装置及电子设备

(57) 摘要

本发明实施例提供一种档案合并方法、装置及电子设备,该方法包括以下步骤:检测多个待合并档案中每两个待合并档案之间的活动规律相识度,待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征;若检测到多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;将档案相似度达到预设相似度阈值的两个目标档案进行合并。这样可以通过两个待合并档案之间的活动规律相识度以及档案相似度两重阈值判断条件,能够很好的解决单纯的基于档案之间的档案相似度进行档案合并效果并不理想的问题。能够大大提高档案合并的效率和准确性。



1. 一种档案合并方法,其特征在于,所述方法包括步骤:

检测多个待合并档案中每两个待合并档案之间的活动规律相识度,所述待合并档案包含用于计算活动规律相识度的多个活动规律特征以及用于计算档案相似度的档案特征,一个所述待合并档案对应于一个人,每个所述待合并档案由不同系统或不同部门建档得到,每两个待合并档案之间的活动规律相识度数量为多个;

若两个待合并档案之间的活动规律相识度中每个活动规律相识度大于预设相识度阈值时,确定检测到所述多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以所述两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;

将档案相似度达到预设相似度阈值的目标档案进行合并;

所述检测多个待合并档案中每两个待合并档案之间的活动规律相识度的步骤包括:

提取每个待合并档案的活动规律特征;

计算每两个活动规律特征之间的活动规律特征相识度;

基于所述每两个活动规律特征之间的活动规律特征相识度,确定多个待合并档案中每两个待合并档案之间的活动规律相识度。

2. 如权利要求1所述的档案合并方法,其特征在于,所述活动规律特征包括多个时间对应的多个活动规律特征值,所述计算每两个活动规律特征之间的活动规律相识度的步骤包括:

计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度;

根据每两个活动规律特征中的至少一个所述活动规律特征值相识度,确定每两个活动规律特征之间的活动规律特征相识度。

3. 如权利要求1所述的档案合并方法,其特征在于,所述计算两个目标档案之间的档案相似度的步骤包括:

提取两个目标档案的档案特征;

计算两个档案特征之间的档案特征相似度;

基于所述两个档案特征之间的档案特征相似度,确定两个目标档案的档案相似度。

4. 如权利要求3所述的档案合并方法,其特征在于,所述档案特征包含多个档案子特征;所述计算两个档案特征之间的档案特征相似度的步骤包括:

根据笛卡尔积策略计算两个档案特征中各个档案子特征之间的档案子特征相似度;

基于两个档案特征中计算得到的各个档案子特征相似度,确定两个档案特征之间的档案特征相似度。

5. 一种档案合并装置,其特征在于,所述装置包括步骤:

检测模块,用于检测多个待合并档案中每两个待合并档案之间的活动规律相识度,所述待合并档案包含用于计算活动规律相识度的多个活动规律特征以及用于计算档案相似度的档案特征,一个所述待合并档案对应于一个人,每个所述待合并档案由不同系统或不同部门建档得到,每两个待合并档案之间的活动规律相识度数量为多个;

计算模块,用于若两个待合并档案之间的活动规律相识度中每个活动规律相识度大于预设相识度阈值时,确定检测到所述多个待合并档案中存在两个待合并档案之间的活动规

律相识度达到预设相识度阈值,则以所述两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;

合并模块,用于将档案相似度达到预设相似度阈值的档案进行合并;

所述检测模块包括:

第一提取单元,用于提取每个待合并档案的活动规律特征;

第一计算单元,用于计算每两个活动规律特征之间的活动规律特征相识度;

第一确定单元,基于所述每两个活动规律特征之间的活动规律特征相识度,确定多个待合并档案中每两个待合并档案之间的活动规律相识度。

6.如权利要求5所述的档案合并装置,其特征在于,所述活动规律特征包括多个时间对应的多个活动规律特征值,所述第一计算单元包括:

第一计算子单元,用于计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度;

第一确定子单元,用于根据每两个活动规律特征中的至少一个所述活动规律特征值相识度,确定每两个活动规律特征之间的活动规律特征相识度。

7.一种电子设备,其特征在于,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至4中任一项所述的档案合并方法中的步骤。

8.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质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如权利要求1至4中任一项所述的档案合并方法中的步骤。

一种档案合并方法、装置及电子设备

技术领域

[0001] 本发明涉及档案管理技术领域,尤其涉及一种档案合并方法、装置及电子设备。

背景技术

[0002] 随着社会的进步,人员流通更加普遍,这样就加大了城市级大规模人员管理的难度。目前有一些部门或者系统可以为每个人员建立人员档案,以方便对人员的管理。其中一种是通过摄像头拍摄人员图像并基于该摄像头下拍摄到的图像建立人员档案。为了更有效的管理人员,一般将摄像头拍摄到的不同的人分别建立对应的人员档案,若某个人再次被同一摄像头拍摄,则将当前拍摄到的图像归入已经建立的人员档案中。但,由于现实生产环境遮挡、角度、光线等问题导致图片在聚类归档中会出现同一个人有多个档案的情况。从而造成多个人有多个档案,使得个人档案的数据量增加。使得在处理城市级大规模个人档案时,档案数量过于庞大,如百亿级的数量。不便于城市级大规模档案管理,需要将同一个人的多个档案进行合并,进而减少个人档案数量。

[0003] 但,单纯的基于档案之间的相识度将同一个人的多个档案进行合并效果并不理想。因为档案相识度阈值设的比较高时重复档案合并率低,当档案相识度阈值设置的低是重复档案就会合并错误。可见现有的档案合并方法中存在档案合并的效率以及准确性不高的问题。

发明内容

[0004] 本发明实施例提供一种档案合并方法,能够提高提高档案合并的效率和准确性。

[0005] 第一方面,本发明实施例提供一种档案合并方法,所述方法包括步骤:

[0006] 检测多个待合并档案中每两个待合并档案之间的活动规律相识度,所述待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征;

[0007] 若检测到所述多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以所述两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;

[0008] 将档案相似度达到预设相似度阈值的目标档案进行合并。

[0009] 可选的,所述检测多个待合并档案中每两个待合并档案之间的活动规律相识度的步骤包括:

[0010] 提取每个待合并档案的活动规律特征;

[0011] 计算每两个活动规律特征之间的活动规律特征相识度;

[0012] 基于所述每两个活动规律特征之间的活动规律特征相识度,确定多个待合并档案中每两个待合并档案之间的活动规律相识度。

[0013] 可选的,所述活动规律特征包括多个时间对应的多个活动规律特征值,所述计算每两个活动规律特征之间的活动规律相识度的步骤包括:

[0014] 计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活

动规律特征值相识度；

[0015] 根据每两个活动规律特征中的至少一个所述活动规律特征值相识度，确定每两个活动规律特征之间的活动规律特征相识度。

[0016] 可选的，所述计算两个目标档案之间的档案相似度的步骤包括：

[0017] 提取两个目标档案的档案特征；

[0018] 计算两个档案特征之间的档案特征相似度；

[0019] 基于所述两个档案特征之间的档案特征相似度，确定两个目标档案的档案相似度。

[0020] 可选的，所述档案特征包含多个档案子特征；所述计算两个档案特征之间的档案特征相似度的步骤包括：

[0021] 根据笛卡尔积策略计算两个档案特征中各个档案子特征之间的档案子特征相似度；

[0022] 基于两个档案特征中计算得到的各个档案子特征相似度，确定两个档案特征之间的档案特征相似度。

[0023] 第二方面，本发明实施例还提供一种档案合并装置，所述方法包括步骤：

[0024] 检测模块，用于检测多个待合并档案中每两个待合并档案之间的活动规律相识度，所述待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征；

[0025] 计算模块，用于若检测到所述多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值，则以所述两个待合并档案为目标档案，并计算两个目标档案之间的档案相似度；

[0026] 合并模块，用于将档案相似度达到预设相似度阈值的目标档案进行合并。

[0027] 可选的，所述检测模块包括：

[0028] 第一提取单元，用于提取每个待合并档案的活动规律特征；

[0029] 第一计算单元，用于计算每两个活动规律特征之间的活动规律特征相识度；

[0030] 第一确定单元，基于所述每两个活动规律特征之间的活动规律特征相识度，确定多个待合并档案中每两个待合并档案之间的活动规律相识度。

[0031] 可选的，所述活动规律特征包括多个时间对应的多个活动规律特征值，所述计算单元包括：

[0032] 第一计算子单元，用于计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度；

[0033] 第一确定子单元，用于根据每两个活动规律特征中的至少一个所述活动规律特征值相识度，确定每两个活动规律特征之间的活动规律特征相识度。

[0034] 第三方面，本发明实施例还提供一种电子设备，包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述实施例提供的档案合并方法中的步骤。

[0035] 第四方面，本发明实施例还提供一种计算机可读存储介质，所述计算机可读存储介质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实施例提供的档案合并方法中的步骤。

[0036] 本发明实施例中,在本发明实施例中,通过检测多个待合并档案中每两个待合并档案之间的活动规律相识度,所述待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征;若检测到所述多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以所述两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;将档案相似度达到预设相似度阈值的档案进行合并。这样可以通过两个待合并档案之间的活动规律相识度以及档案相似度两重阈值判断条件,能够很好的解决单纯的基于档案之间的档案相似度进行档案合并效果并不理想的问题。因为同一个人的活动规律每天的活动规律基本上都是一致的,所以如果是同一人的档案那么这两个档案之间的活动规律基本相同。然后在两个待合并档案的活动规律相识度满足预设相识度的基础上,再计算他们之间的档案相似度,这样能大大提高档案合并的效率和准确性。

附图说明

[0037] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0038] 图1是本发明实施例提供的一种档案合并方法的流程图;

[0039] 图2是图1实施例中步骤101提供的一种方法的流程图;

[0040] 图3是图2实施例中步骤202提供的一种方法的流程图;

[0041] 图4是图1实施例中步骤102提供的一种方法的流程图;

[0042] 图5是图4实施例中步骤402提供的一种方法的流程图;

[0043] 图6是本发明实施例提供的一种档案合并装置的结构示意图;

[0044] 图7是图6实施例中检测模块提供的一种结构示意图;

[0045] 图8是图7实施例中第一计算单元提供的一种结构示意图;

[0046] 图9是图6实施例中计算模块提供的一种结构示意图;

[0047] 图10是图9实施例中第二计算单元提供的一种结构示意图;

[0048] 图11是本发明实施例提供的一种电子设备的结构示意图。

具体实施方式

[0049] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0050] 请参见图1,图1是本发明实施例提供的一种档案合并方法的流程图,方法包括步骤:

[0051] 步骤101、检测多个待合并档案中每两个待合并档案之间的活动规律相识度。

[0052] 其中,待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征。

[0053] 上述待合并档案为需要合并的同一个人的多个档案。一个档案对应的是同一个人。当然了,若想合并多个人的多个档案时,该待合并档案为多个人的多个档案。待合并档案可以为图片档案,以图片进行归档,当然了,待合并档案还可以是文字档案或属性档案等。在本实施例中主要以图片档案为主进行说明。需要说明的是,这里所说的图片也可以是图像、照片等。

[0054] 上述活动规律为每个待合并档案中某个人的活动轨迹规律。每个待合并档案中的图片按时间戳进行排序,进而可以体现某个人的活动轨迹规律。每个待合并档案均有自己的活动规律,当两个待合并档案均为同一个人的档案时,这两个待合并档案中体现的活动规律基本是一致的,因为同一个人每天的活动规律基本上都是一致的。

[0055] 上述活动规律相识度则为每个待合并档案之间的活动规律之间的相识度。上述活动规律特征为能够代表该活动规律的特征,比如,活动轨迹规律的活动地点特征等。可以通过该活动规律特征来计算得到该活动规律的活动规律相识度。

[0056] 上述档案相似度为两个待合并档案之间的相似度。上述档案特征可以为能够代表待合并档案的特征,比如,图片档案中每个档案封面的封面特征,当档案封面为人脸封面时,档案特征则为人脸特征,档案相似度则为两个人脸特征之间的相似度。属性档案中每个档案属性的属性特征。档案属性可以是某个人的指纹、DNA等,还可以是身高、体重、体型等。属性特征可以包含有人员对应的指纹特征、DNA特征等。

[0057] 具体的,当获取到多个待合并档案时,需要计算多个待合并档案中任意两个待合并档案的活动规律之间的相识度。

[0058] 需要说明的是,多个待合并档案可以存储在档案数据库中或者存在档案云空间中,还可以存储在档案系统中,当然了,待合并档案的数量随着时间的增加而增加。且每个待合并档案可以是由不同系统或不同部门建档得到的,例如,在某小区管理系统中的摄像头拍摄到某人员的人脸图像,该系统基于拍摄到的人脸图像建立该人员的档案,并存储在对应的档案数据库中;而在另一小区管理系统中的摄像头也拍摄到该人员的人脸图像,也对该人员建立对应的档案,也存储在对应的档案数据库。此时,该人员在两个小区管理系统的档案数据库中都存有档案,也就是说该人员同时有两个档案,但这两个档案均是同一个人的。

[0059] 或者,当某个摄像头拍摄到某个人的人脸图像并建立个人图片档案后,该摄像头再次拍摄到该人员的人脸图像时,由于现实生产环境遮挡、角度、光线等问题使得摄像头认为不是同一个人,所以重新为该人员重新建立一个新的图片档案,这样该人员也就有两个档案,并存储到对应的档案数据库中。

[0060] 这样,当需要对多个小区管理系统的人员档案进行管理时,需要将多个小区管理系统中的档案数据库中的档案拿出来进行管理,这样就会存在同一个人有多个档案的情况。并且每个小区管理系统都会建立多个人员的档案,那么管理多个小区管理系统的档案时,就会存在多个人对应多个档案的情况。为此,上述多个待合并档案可以相当于本例子中,多个小区管理系统中提供的多个档案,这多个档案可以是多个小区管理系统提供的部分档案数据或者全部档案数据。每个待合并档案还可以是相同系统或相同部门的不同建档终端建档得到的。

[0061] 步骤102、若检测到多个待合并档案中存在两个待合并档案之间的活动规律相识

度达到预设相识度阈值,则以两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度。

[0062] 其中,上述预设相识度阈值可以是预先设置的相识度门限值。该预设相识度阈值是用于判断两个待合并档案中的活动规律之间是否一致的标准,该预设相识度阈值可以根据需要设置。两个待合并档案的活动规律相识度越大,说明两个待合并档案中的活动规律越一致,当然了,在设置预设相识度阈值的情况下,当两个待合并档案的活动规律相识度满足预设相识度阈值时,才能说明这两个待合并档案之间的活动规律基本一致,那么这两个待合并档案则为目标档案。

[0063] 上述两个待合并档案为目标档案是指这两个待合并档案为需要计算档案相似度的待合并档案。

[0064] 具体的,当检测到多个待合并档案中每两个待合并档案之间的活动规律相识度后,将检测到的每个活动规律相识度与预设相识度阈值进行比较判断。活动规律相识度的数量可以是一个或多个。当检测到一个或多个活动规律相识度时,均将每个活动规律相识度与预设相识度阈值进行比较,并判断每个活动规律相识度是否满足预设相识度阈值。若每个活动规律相识度大于预设相识度阈值,说明该活动规律相识度满足预设相识度阈值,也即说明该活动规律相识度对应的两个待合并档案的活动规律基本一致,则将这两个待合并档案作为目标档案保留下来,并计算这两个待合并档案的档案相似度。

[0065] 当然,若每活动规律相识度小于等于预设相识度阈值,说明该活动规律相识度不满足预设相识度阈值,也即说明该活动规律相识度对应的两个待合并档案的活动规律不一致,则不需要将这两个待合并档案作为目标档案保留下来,并且也不需要计算这两个待合并档案之间的档案相似度。

[0066] 需要说明的是,当计算得到多个活动规律相识度均满足预设相识度时,对应的得到多组目标档案。并分别计算多组目标档案的档案相似度。也就说明档案相似度数量也为多个。

[0067] 步骤103、将档案相似度达到预设相似度阈值的两个目标档案进行合并。

[0068] 其中,上述预设相似度阈值可以是预先设置的相似度门限值。该预设相似度阈值是用于判断两个待合并档案是否是同一个人的档案的条件。该预设相似度阈值可以根据需要设置,比如,设置相似度门限值为0.90,如果两个待合并档案之间的相似度为0.95,说明两个待合并档案的档案相似度0.95大于预设相似度阈值0.90,则认为这两个待合并档案是同一个人的档案。上述合并是指将两个待合并档案融合为一个档案,也就是将属于同一个人的两个待合并档案合并为一个档案。

[0069] 具体的,当计算得到两个目标档案之间的档案相似度后,将计算得到的档案相似度与预设相似度阈值进行比较判断,并判断档案相似度是否满足预设相似度阈值。若档案相似度大于预设相似度阈值,说明该档案相似度满足预设相似度阈值,也即说明该档案相似度对应的两个目标档案对应的两个待合并档案为同一个人的档案,则将这两个待合并档案进行档案合并,得到一个档案,这样可以将属于同一个人的档案合为一个档案。

[0070] 当然,若档案相似度小于或等于预设相似度阈值,说明该档案相似度不满足预设相似度阈值,也即说明该档案相似度对应的两个目标档案对应的两个待合并档案不是同一个人的档案,则不将这两个待合并档案进行档案合并。

[0071] 进一步的,当多个待合并档案中一个人同时存在多个档案时,可以将多个同时满足活动规律相识度以及档案相似度的档案共同合并为一个档案。

[0072] 在本发明实施例中,通过检测多个待合并档案中每两个待合并档案之间的活动规律相识度,待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征;若检测到多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度;将档案相似度达到预设相似度阈值的目标档案进行合并。这样可以通过两个待合并档案之间的活动规律相识度以及档案相似度两重阈值判断条件,能够很好的解决单纯的基于档案之间的档案相似度进行档案合并效果并不理想的问题。因为同一个人的活动规律每天的活动规律基本上都是一致的,所以如果是同一人的档案那么这两个档案之间的活动规律基本相同。然后在两个待合并档案的活动规律相识度满足预设相识度的基础上,再计算他们之间的档案相似度,这样能大大提高档案合并的效率和准确性。

[0073] 参见图2,图2是图1实施例中步骤101提供的一种方法的流程图,步骤101包括步骤:

[0074] 步骤201、提取每个待合并档案的活动规律特征。

[0075] 步骤202、计算每两个活动规律特征之间的活动规律特征相识度。

[0076] 步骤203、基于每两个活动规律特征之间的活动规律特征相识度,得到多个待合并档案中每两个待合并档案之间的活动规律相识度。

[0077] 上述活动规律特征相识度为两个待合并档案中活动规律的活动规律特征之间的相识度。该活动规律特征相识度是用于判断两个判断两个活动规律之间是否一致的条件。

[0078] 具体的,首先提取多个待合并档案中,每个待合并档案的活动规律特征。然后再分别计算每两个待合并档案的活动规律中对应的活动规律特征之间的相识度。并且将这两个活动规律特征对应的活动规律特征相识度作为两个活动规律之间的规律相识度,进而得到两个待合并档案之间的活动规律相识度。

[0079] 进一步的,参见图3,图3是图2实施例中步骤202提供的一种方法的流程图,活动规律特征包括多个时间对应的多个活动规律特征值,步骤202包括:

[0080] 步骤301、计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度。

[0081] 步骤302、根据每两个活动规律特征中的至少一个活动规律特征值相识度,得到每两个活动规律特征之间的活动规律特征相识度。

[0082] 其中,上述时间可以是按年、月、日、时来进行设置。上述相同时间可以是相同年、相同月、相同日、相同时等。上述活动规律特征值为活动规律特征中某个时间下的活动规律特征对应的特征值。上述活动规律特征值的数量可以为一个或多个。对应的,上述活动规律特征值相识度的数量也可以为一个或多个,当两个活动规律特征存在多个相同时间下的两个活动规律特征值时,活动规律特征值相识度也为多个。

[0083] 具体的,以档案A、A1、B、C各个日期下的活动规律特征值(该活动规律特征值设置为24维的整型数组,进而可以代表24个小时对应的活动规律特征值)为例进行说明,提供的档案A、A1、B、C的活动规律特征值分别如下:

[0084] A(1号):[1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];

- [0085] A(2号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0086] A(3号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0087] A(4号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0088] A(5号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]。
- [0089] A1(1号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0090] A1(3号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0091] A1(4号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0092] A1(5号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0093] A1(7号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]。
- [0094] B(1号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0095] B(4号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0096] B(6号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0097] B(7号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0098] B(11号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]。
- [0099] C(3号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0100] C(5号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0101] C(8号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0102] C(12号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯];
- [0103] C(23号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]。

[0104] 其中,A(1号)、A(2号)等均表示为A档案的1号和2号的活动规律特征。当然了,A(1号)、A(2号)等中的日期号数可以根据实际情况进行确定。A(1号): [1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]表示为A档案在1号的活动规律特征中24小时对应的活动规律特征值。在本实施例中,[1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]中的数值则为24小时对应的活动规律特征值的具体数值,当然,[1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]中的数值仅是示例性的,具体的数值大小可根据实际的活动规律特征值进行设置。当然了,在计算各个活动规律特征值之间的活动规律特征值时,可将[1,2,3,4,5,6,7,8,9,10,11,12,13,⋯]中的数值转换为二进制数组,进而根据二进制数进行计算。例如,假设A(1号)的活动规律特征值为:[157(1001 1101),78(0100 1110),52(0011 0100),51(0011 0011),⋯]

[0105] 需要说明的是,在本实施例中,每个待合并档案中的活动规律以摄像头映射为基础。具体的,可以设置摄像头编号以及摄像头编号对应的摄像头编码,其中,摄像头编号设置为:camera8、camera7、camera6、camera5、camera4、camera3、camera2、camera1;摄像头编码设置为:1、1、1、1、1、1、1、1。

[0106] 所以可以用一个整型255表示,因为255的二进制是11111111,其中每一位都代表一个摄像头。在具体实施时,当某个人出现在摄像头中时则摄像头编码为1,当某个人没有出现在摄像头中时,则摄像头编码为0。比如,档案A中有一活动规律特征值为157(1001 1101)时,则说明档案A中的这个人出现在camera 1、camera 3、camera 4、camera 5、camera 8中。

[0107] 进一步的,在获取到档案A、A1、B、C中每个档案在各个时间下的活动规律特征值后,分别对档案A、A1、B、C按时间取交集得到以下活动规律特征值,以档案A和档案A1按时间

取交集为例,具体如下:

[0108] $A \cap A1$:

[0109] A(1号)A(2号)A(3号)A(4号)A(5号)

[0110] \cap

[0111] A1(1号)A1(3号)A1(4号)A1(5号)A1(7号)

[0112] =A(1号)A(3号)A(4号)A(5号)和A1(1号)A1(3号)A1(4号)A1(5号)。

[0113] 又进一步的,假设A(1号)特征值为:[157(1001 1101),78(0100 1110),52(0011 0100),51(0011 0011),...]

[0114] 假设A1(1号)特征值为:[156(1001 1100),76(0100 1100),52(0011 0100),50(0011 0000),...]。

[0115] 所以A(1号)_A1(1号) = (二进制每位数字相同的个数)/(二进制所有位数)之和

[0116] = (157(1001 1101)和156(1001 1100)为7)7/8

[0117] +(78(0100 1110)和76(0100 1100)为7)7/8

[0118] +(52(0011 0100)和52(0011 0100)为8)8/8

[0119] +(51(0011 0011)和50(0011 0000)为6)6/8

[0120] +...

[0121] =7/8+7/8+8/8+6/8+... (一共是24维的数据)

[0122] =19.8。

[0123] 最后再用 $19.8/24=0.825$,所以最终计算得到的档案A(1号)_A1(1号)的活动规律特征值相识度为0.825。

[0124] 同理得到:

[0125] A(3号)_A1(3号)=0.86;

[0126] A(4号)_A1(4号)=0.88;

[0127] A(5号)_A1(5号)=0.90。

[0128] 最终得到档案A_A1的活动规律相识度为: $0.825+0.86+0.88+0.90=3.465$ 。

[0129] 同理得到档案A_B,档案A_C,档案A1_B,档案A1_C,档案B_C的活动规律相识度为:

[0130] A_A1=3.456;

[0131] A_B=1.25;

[0132] A_C=0.105;

[0133] A1_B=1.9;

[0134] A1_C=1.2;

[0135] B_C=2.8。

[0136] 将计算得到活动规律特征值相识度确定为两个活动规律特征之间的活动规律特征相识度,进而得到两个活动规律之间的活动规律相识度,也进而得到两个待合并档案之间的活动规律相识度。

[0137] 若预设的预设相识度阈值为2,则在档案A、A1、B、C四个档案中,得到档案A_A1以及档案B_C的活动规律相识度满足预设相识度阈值,则说明档案A,档案A1之间的活动规律基本一致,档案B和档案C之间的活动规律基本一致。而A_B=1.25;A_C=0.105;A1_B=1.9;A1_C=1.2不满足预设相识度阈值,且互相之间的活动规律不一致。

[0138] 在本实施例中,可以根据两个待合并档案中的活动规律特征以及对应的不同时间下的活动规律特征值来计算两个待合并档案中对应的人员的活动规律的相识度,进而判断出两个待合并档案对应的人员的活动规律是否一致,若一致则说明这两个待合并档案可能是同一个人的档案,则可以执行后续步骤。若不一致,则说明这两个待合并档案一定不是同一个人的档案,不需要执行后续步骤。

[0139] 参见图4,图4是图1实施例中步骤102提供的一种方法的流程图。

[0140] 步骤102包括步骤:

[0141] 步骤401、提取两个目标档案的档案特征。

[0142] 步骤402、计算两个档案特征之间的档案特征相似度。

[0143] 步骤403、基于两个档案特征之间的档案特征相似度,得到两个待合并档案的档案相似度。

[0144] 上述档案特征相似度为两个目标档案中的档案特征之间的档案特征相似度。档案特征相似度是用于判断两个目标档案是否为一个人员的档案的条件。

[0145] 具体的,当确定两个待合并档案之间的活动规律相识度满足预设相识度后,也就确定这两个待合并档案为目标档案,并需要计算这两个待合并档案之间的档案相似度。为此,需要提取这两个目标档案的档案特征,并计算两个档案特征之间的档案特征相似度,并将这个档案特征相似度确定为这两个目标档案之间的档案相似度,因为档案特征能够代表目标档案。比如,档案特征为档案封面特征,则提取这两个目标档案对应的档案封面特征,并计算两个档案封面特征之间的档案封面特征相似度,进而得到两个目标档案的档案相似度。

[0146] 进一步的,参见图5,图5是图4实施例中步骤402提供的一种方法的流程图。档案特征包含多个档案子特征,步骤402包括步骤:

[0147] 步骤501、根据笛卡尔积策略计算两个档案特征中各个档案子特征之间的档案子特征相似度。

[0148] 步骤502、基于两个档案特征中计算得到的各个档案子特征相似度,得到两个档案特征之间的档案特征相似度。

[0149] 其中,上述笛卡尔积又称为笛卡尔乘积,笛卡尔乘积是指在数学中,两个集合X和Y的笛卡尔积(Cartesian product),又称直积,表示为 $X \times Y$,第一个对象是X的成员而第二个对象是Y的所有可能有序对的其中一个成员,假设集合 $A = \{a, b\}$,集合 $B = \{0, 1, 2\}$,则两个集合的笛卡尔积为 $\{(a, 0), (a, 1), (a, 2), (b, 0), (b, 1), (b, 2)\}$ 。上述笛卡尔积策略即为采用笛卡尔积计算各个档案子特征之间的相似度的方法。

[0150] 上述档案子特征可以为每个档案特征中的多个特征,比如,档案特征为档案封面特征时,若待合并档案中的档案封面设置有多多个封面代表图片时,一个档案子特征则为一个封面代表图片对应的特征。

[0151] 上述档案子特征相似度为每个档案子特征之间的相似度,是用于判断各个档案子特征之间是否为同一个档案子特征的条件。档案子特征相似度的数量可以为一个或多个。当待合并档案均只设置有一个档案子特征时,得到的档案子特征相似度也只为一个,当待合并档案均设置有多多个档案子特征时,得到的档案子特征相似度也只为多个。

[0152] 具体的,通过笛卡尔积策略计算得到目标档案中档案子特征之间的档案子特征相

似度。并且可以计算各个档案子特征相似度的平均相似度,进而得到两个目标档案之间的档案相似度。

[0153] 示例性的,假如档案E有e1档案子特征、e2档案子特征、e3档案子特征;档案F有f1档案子特征、f2档案子特征、f3档案子特征。那么可以根据笛卡尔积策略计算得到档案A和档案B之间的相似度,例如:

[0154] $e1_f1=0.91, e1_f2=0.94, e1_f3=0.93;$

[0155] $e2_f1=0.93, e2_f2=0.91, e2_f3=0.90;$

[0156] $e3_f1=0.95, e3_f2=0.90, e3_f3=0.96。$

[0157] 求得他们的平均值 $= (0.91+0.94+0.93+0.93+0.91+0.90+0.95+0.90+0.96)/9=0.926。$

[0158] 具体的,假如通过上述笛卡尔积策略可以得到档案A_A1和B_C档案相似度为:A_A1:0.818;B_C:0.77。若设置预设相似度阈值为0.816,其中A_A1:0.818大于预设相似度阈值,所以档案A和档案A1是同一个人的档案。相反档案B和档案C不是同一个人的档案。

[0159] 在本实施例中,通过计算两个目标档案中的档案特征对应的档案子特征之间的档案子特征相似度,进而得到两个档案特征之间的档案特征相似度,进而得到两个目标档案之间的档案相似度。进而通过通过计算两个待合并档案之间的活动规律相识度以及档案相似度来判断这两个待合并档案是否为同一个人的档案。若为同一个人的档案,则将这两个待合并档案合并为一个档案。进而提高档案合并的效率以及准确性。

[0160] 参见图6,图6是本发明实施例提供的一种档案合并装置的结构示意图,档案合并装置600包括步骤:

[0161] 检测模块601,用于检测多个待合并档案中每两个待合并档案之间的活动规律相识度,待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征。

[0162] 计算模块602,用于若检测到多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度。

[0163] 合并模块603,用于将档案相似度达到预设相似度阈值的目标档案进行合并。

[0164] 参见图7,图7是图6实施例中检测模块提供的一种结构示意图,检测模块601包括:

[0165] 第一提取单元6011,用于提取每个待合并档案的活动规律特征。

[0166] 第一计算单元6012,用于计算每两个活动规律特征之间的活动规律特征相识度。

[0167] 第一确定单元6013,基于每两个活动规律特征之间的活动规律特征相识度,确定多个待合并档案中每两个待合并档案之间的活动规律相识度。

[0168] 参见图8,图8是图7实施例中第一计算单元提供的一种结构示意图,活动规律特征包括多个时间对应的多个活动规律特征值,第一计算单元6012包括:

[0169] 第一计算子单元60121,用于计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度。

[0170] 第一确定子单元60122,用于根据每两个活动规律特征中的至少一个活动规律特征值相识度,确定每两个活动规律特征之间的活动规律特征相识度。

[0171] 参见图9,图9是图6实施例中计算模块提供的一种结构示意图,计算模块602包括:

- [0172] 第二提取单元6021,用于提取两个目标档案的档案特征。
- [0173] 第二计算单元6022,用于计算两个档案特征之间的档案特征相似度。
- [0174] 第二确定单元6023,用于基于两个档案特征之间的档案特征相似度,确定两个目标档案的档案相似度。
- [0175] 参见图10,图10是图9实施例中第二计算单元提供的一种结构示意图,档案特征包含多个档案子特征;第二计算单元6022包括:
- [0176] 第二计算子单元60221,用于根据笛卡尔积策略计算两个档案特征中各个档案子特征之间的档案子特征相似度。
- [0177] 第二确定子单元60222,用于基于两个档案特征中计算得到的各个档案子特征相似度,确定两个档案特征之间的档案特征相似度。
- [0178] 本发明实施例提供的档案合并装置600能够实现上述方法实施例中的各个实施方式,以及相应有益效果,为避免重复,这里不再赘述。
- [0179] 参见图11,图11是本发明实施例提供的一种电子设备700的结构示意图,该电子设备700包括:存储器702、处理器701及存储在存储器702上并可在处理器701上运行的计算机程序,处理器701执行计算机程序时实现上述实施例提供的档案合并方法中的步骤,处理器701执行以下步骤:
- [0180] 检测多个待合并档案中每两个待合并档案之间的活动规律相识度,待合并档案包含用于计算活动规律相识度的活动规律特征以及用于计算档案相似度的档案特征。
- [0181] 若检测到多个待合并档案中存在两个待合并档案之间的活动规律相识度达到预设相识度阈值,则以两个待合并档案为目标档案,并计算两个目标档案之间的档案相似度。
- [0182] 将档案相似度达到预设相似度阈值的目标档案进行合并。
- [0183] 可选的,处理器701执行的检测多个待合并档案中每两个待合并档案之间的活动规律相识度的步骤包括:
- [0184] 提取每个待合并档案的活动规律特征。
- [0185] 计算每两个活动规律特征之间的活动规律特征相识度。
- [0186] 基于每两个活动规律特征之间的活动规律特征相识度,确定多个待合并档案中每两个待合并档案之间的活动规律相识度。
- [0187] 可选的,活动规律特征包括多个时间对应的多个活动规律特征值,处理器701执行的计算每两个活动规律特征之间的活动规律相识度的步骤包括:
- [0188] 计算每两个活动规律特征中每两个相同时间下的两个活动规律特征值之间的活动规律特征值相识度;
- [0189] 根据每两个活动规律特征中的至少一个活动规律特征值相识度,确定每两个活动规律特征之间的活动规律特征相识度。
- [0190] 可选的,处理器701执行的计算两个目标档案之间的档案相似度的步骤包括:
- [0191] 提取两个目标档案的档案特征。
- [0192] 计算两个档案特征之间的档案特征相似度。
- [0193] 基于两个档案特征之间的档案特征相似度,确定两个目标档案的档案相似度。
- [0194] 可选的,档案特征包含多个档案子特征;处理器701执行的计算两个档案特征之间的档案特征相似度的步骤包括:

[0195] 根据笛卡尔积策略计算两个档案特征中各个档案子特征之间的档案子特征相似度。

[0196] 基于两个档案特征中计算得到的各个档案子特征相似度,确定两个档案特征之间的档案特征相似度。

[0197] 需要说明的是,上述电子设备700可以是档案相关部门的智能终端、手机、平板电脑等设备。

[0198] 本发明实施例提供的电子设备700能够实现上述方法实施例中的各个实施方式,以及相应有益效果,为避免重复,这里不再赘述。

[0199] 本发明实施例还提供一种计算机可读存储介质,计算机可读存储介质上存储有计算机程序,该计算机程序被处理器701执行时实现本发明实施例提供的档案合并方法的各个过程,且能达到相同的技术效果,为避免重复,这里不再赘述。

[0200] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的程序可存储于一计算机可读取存储介质中,该程序在执行时,可包括如上述各方法的实施例的流程。其中,所述的存储介质可为磁碟、光盘、只读存储记忆体(Read-Only Memory,ROM)或随机存取存储器(Random Access Memory,简称RAM)等。

[0201] 以上所揭露的仅为本发明较佳实施例而已,当然不能以此来限定本发明之权利范围,因此依本发明权利要求所作的等同变化,仍属本发明所涵盖的范围。

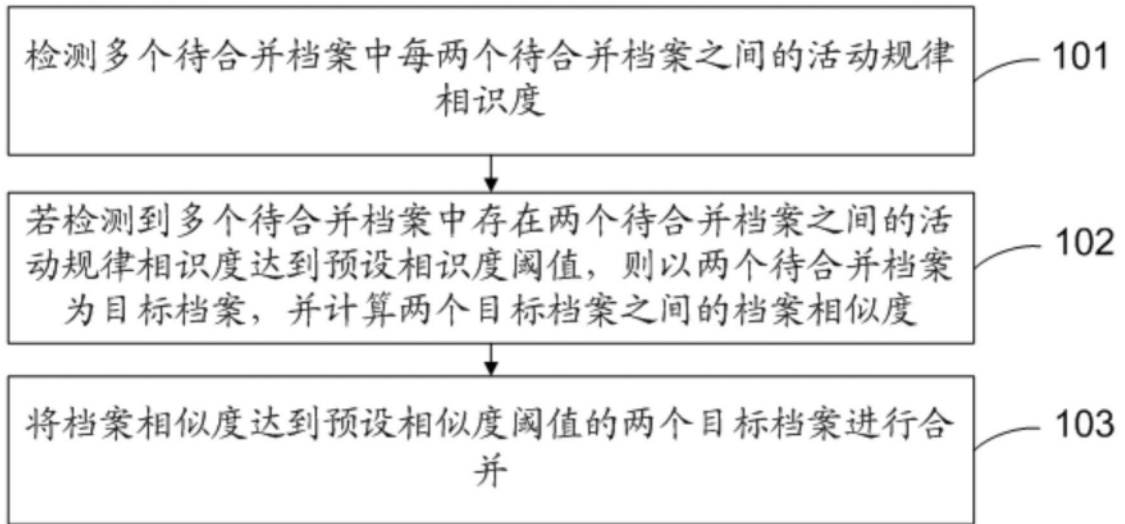


图1

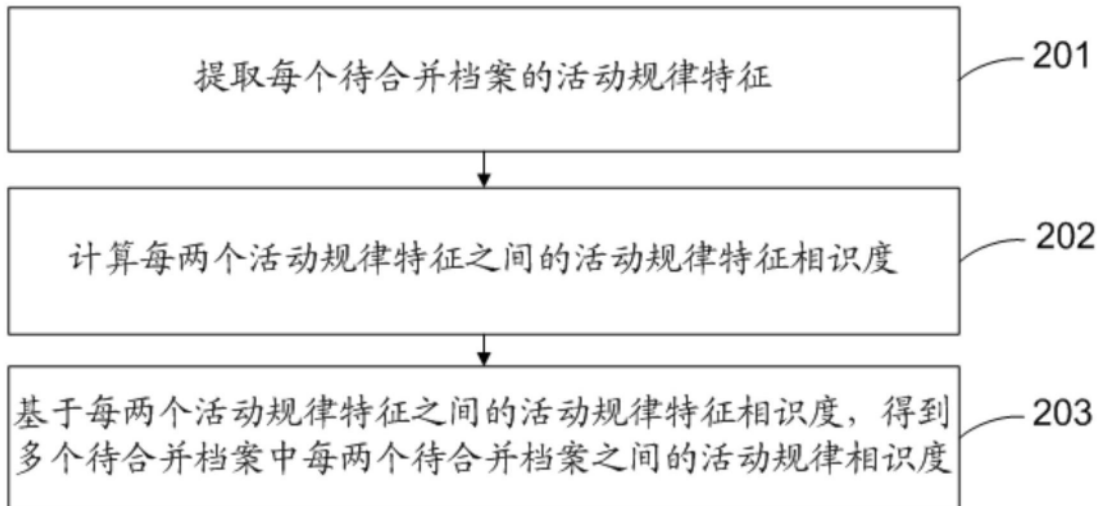


图2

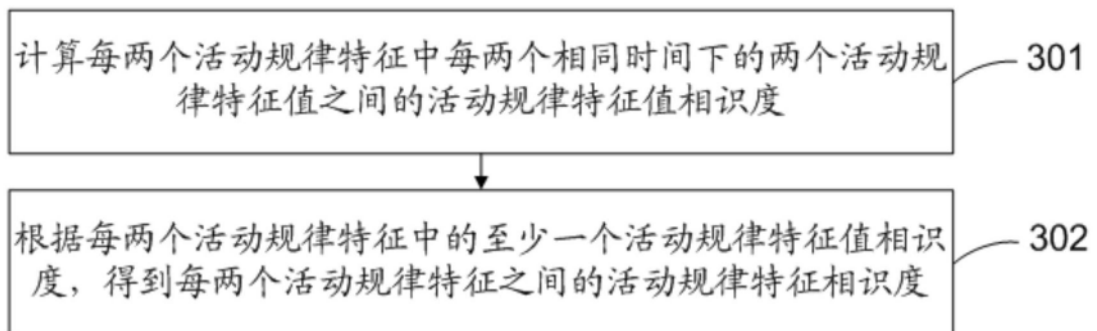


图3

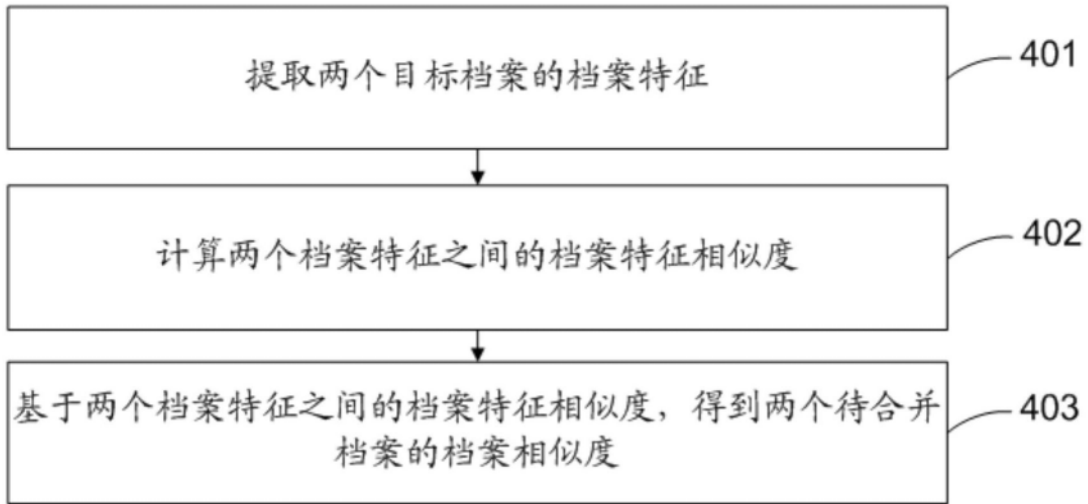


图4

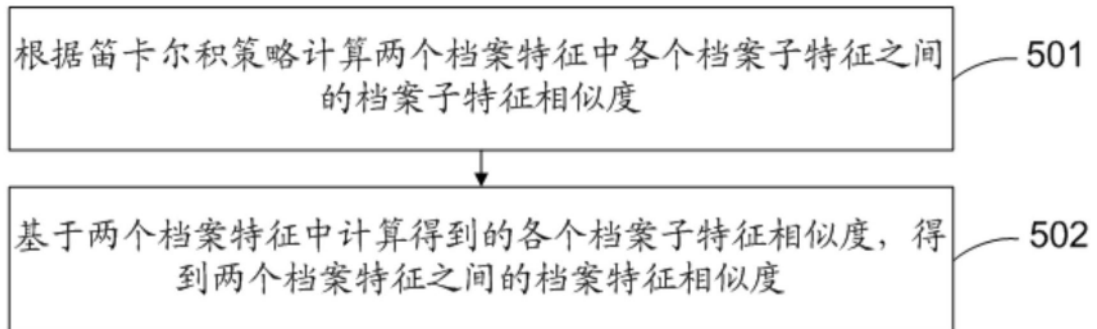


图5

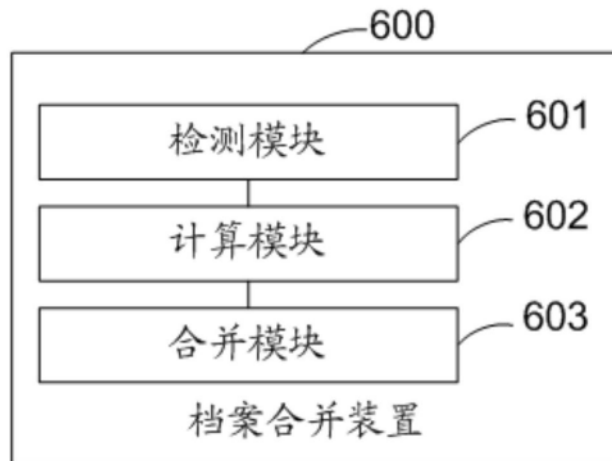


图6

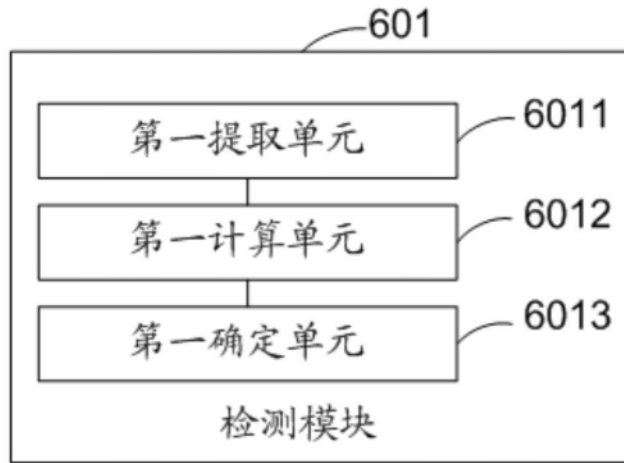


图7

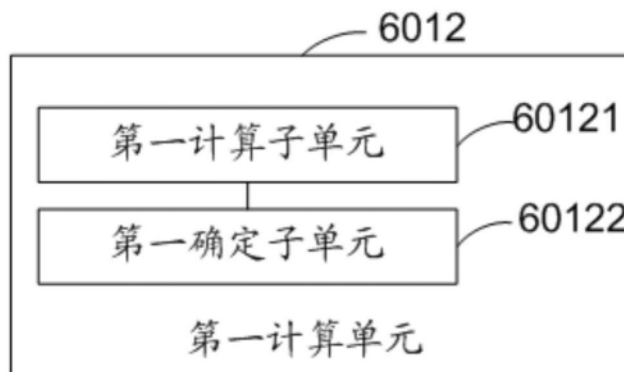


图8

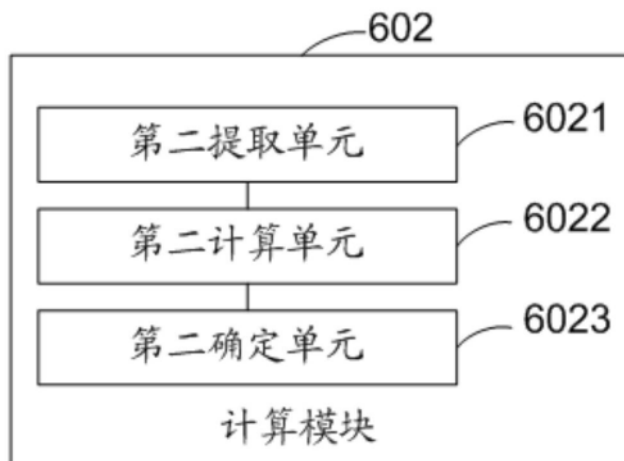


图9

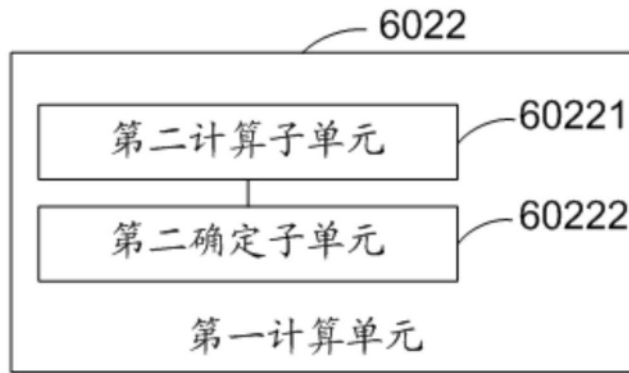


图10

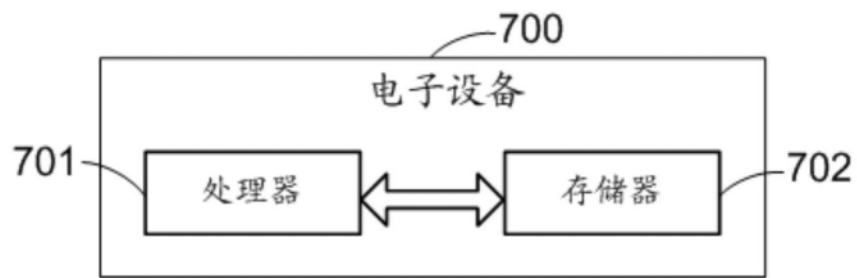


图11