



(12) 发明专利

(10) 授权公告号 CN 110309336 B

(45) 授权公告日 2023. 08. 08

(21) 申请号 201810200560.X

G06F 18/213 (2023.01)

(22) 申请日 2018.03.12

G06F 18/22 (2023.01)

(65) 同一申请的已公布的文献号  
申请公布号 CN 110309336 A

(56) 对比文件

CN 103678702 A, 2014.03.26

CN 106445939 A, 2017.02.22

(43) 申请公布日 2019.10.08

US 8611617 B1, 2013.12.17

(73) 专利权人 腾讯科技(深圳)有限公司  
地址 518000 广东省深圳市南山区高新区  
科技中一路腾讯大厦35层

审查员 孔昕

(72) 发明人 黄桂洲 宋翔宇 贺伟 郭德安  
江启泉

(74) 专利代理机构 深圳市深佳知识产权代理事  
务所(普通合伙) 44285  
专利代理师 王仲凯

(51) Int. Cl.

G06F 16/583 (2019.01)

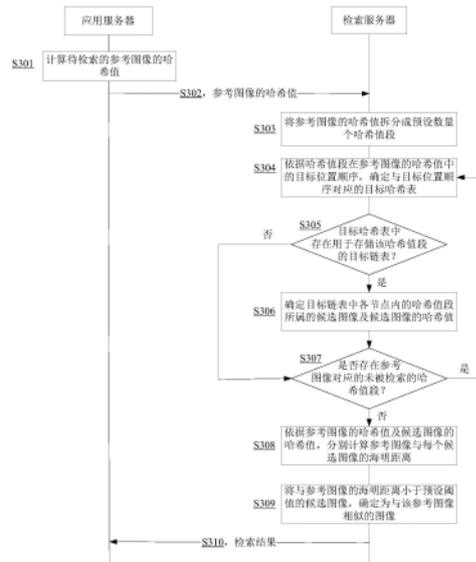
权利要求书5页 说明书17页 附图7页

(54) 发明名称

图像检索方法、装置、系统、服务器以及存储  
介质

(57) 摘要

本申请公开了一种图像检索方法、装置、系  
统、服务器以及存储介质,该方法包括:获取待检  
索的参考图像的第一哈希值;将该第一哈希值拆  
分成预设数量个连续的第一哈希值段,其中,不  
同的第一哈希值段在该第一哈希值中的位置顺  
序不同;根据该第一哈希值段的位置顺序,定位  
目标哈希表,并从该目标哈希表的多个哈希值段  
中,检索出与该第一哈希值段相同的第二哈希值  
段,所述目标哈希表用于存储图像的哈希值中,  
处于所述第一哈希值段的位置顺序上的哈希值  
段;从该第二哈希值段所属的候选图像中,检索  
与该参考图像相似的目标图像。本申请的方案可  
以减少图像检索过程中的数据处理量,提高图像  
检索效率。



CN 110309336 B

1. 一种图像检索方法,其特征在于,包括:

获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像;

所述目标哈希表中存储有至少两条链表,所述链表中的节点用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段,且属于同一个链表的节点中存储的哈希值段相同;所述目标哈希表中的链表通过如下方式生成:

获取数据库中待处理的图像的哈希值;

将所述图像的哈希值拆分为所述预设数量个哈希值段,其中,不同的所述哈希值段在所述图像的哈希值中的位置顺序不同;

对于拆分出的每个哈希值段,定位与所述哈希值段的位置顺序对应的目标哈希表,并检测所述目标哈希表中是否存在用于记录所述哈希值段的链表;

在所述目标哈希表中存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成的节点添加到所述用于记录所述哈希值段的链表的尾端;

在所述目标哈希表中不存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成节点设定为所述目标哈希表中新增的链表的首个节点。

2. 根据权利要求1所述的图像检索方法,其特征在于,所述从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像,包括:

获取所述第二哈希值段所属的候选图像的第二哈希值,所述第二哈希值用于表征候选图像的图像特征;

依据所述参考图像的第一哈希值和所述候选图像的第二哈希值,计算所述候选图像与参考图像的海明距离;

从所述第二哈希值段所属的候选图像中,检索出与所述参考图像的海明距离小于预设阈值的目标图像。

3. 根据权利要求1所述的图像检索方法,其特征在于,所述从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,包括:

从所述目标哈希表的至少两条链表中,定位出目标链表,所述目标链表的各个节点中存储的第二哈希值段与所述第一哈希值段相同。

4. 根据权利要求3所述的图像检索方法,其特征在于,所述链表的节点中还存储有所述节点内的哈希值段所属的图像的图像信息;

在所述从所述目标哈希表的至少两条链表中,定位出用于存储所述第一哈希值段的目标链表之后,还包括:

依据目标链表的节点中的图像信息,获取所述目标链表中每个节点内的第二哈希值所属的候选图像。

5. 根据权利要求1或2所述的图像检索方法,其特征在于,所述根据所述第一哈希值段的位置顺序,定位目标哈希表,包括:

将所述预设数量个连续的第一哈希值段发送给多台从服务器,以便所述从服务器依据所述第一哈希值段的位置顺序,从所述从服务器负责检索的区域段内查询目标哈希表,所述区域段为数据库的存储空间中的部分存储区域;

所述从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,包括:

通过所述从服务器从所述目标哈希表的多个哈希段中,检索出与所述第一哈希值段相同的第二哈希值段;

所述从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像,包括:

接收所述从服务器返回的查询结果,所述查询结果至少包括:从所述第二哈希值段所属的候选图像中,检索出的与所述参考图像相似的目标图像的信息。

6. 根据权利要求5所述的图像检索方法,其特征在于,还包括:

在监测到所述多台从服务器中存在运行异常的故障从服务器的情况下,确定所述故障从服务器当前负责检索的区域段;

将所述故障从服务器对应的区域段分配给至少一个未出现运行异常的从服务器。

7. 一种图像检索装置,其特征在于,包括:

哈希获取单元,用于获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

哈希拆分单元,用于将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

哈希检索单元,根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;所述目标哈希表中存储有至少两条链表,所述链表中的节点用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段,且属于同一个链表的节点中存储的哈希值段相同;

图像匹配单元,用于从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像;

链表生成单元,用于通过如下方式生成所述目标哈希表中的链表:

获取数据库中待处理的图像的哈希值;

将所述图像的哈希值拆分所述预设数量个哈希值段,并获取每个哈希值段在所述图像的哈希值中的位置顺序;

对于拆分出每个哈希值段,定位与所述哈希值段的位置顺序对应的目标哈希表,并检测所述目标哈希表中是否存在用于记录所述哈希值段的链表;

在所述目标哈希表中存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成的节点添加到所述用于记录所述哈希值段的链表的尾端;

在所述目标哈希表中不存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成节点设定为所述目标哈希表中新增链表的首个节点。

8. 根据权利要求7所述的图像检索装置,其特征在于,所述图像匹配单元,包括:

哈希确定单元,用于获取所述第二哈希值段所属的候选图像的第二哈希值,所述第二哈希值用于表征候选图像的图像特征;

距离计算单元,用于依据所述参考图像的第一哈希值和所述候选图像的第二哈希值,计算所述候选图像与参考图像的海明距离;

距离匹配单元,用于从所述第二哈希值段所属的候选图像中,检索出与所述参考图像的海明距离小于预设阈值的目标图像。

9. 根据权利要求7所述的图像检索装置,其特征在于,所述哈希检索单元,包括:

链表定位子单元,用于从所述目标哈希表的至少两条链表中,定位出用于存储所述第一哈希值段的目标链表,所述目标链表的各个节点中存储的第二哈希值段与所述第一哈希值段相同。

10. 根据权利要求9所述的图像检索装置,其特征在于,所述链表的节点中还存储有所述节点内的哈希值段所属的图像的图像信息;所述装置还包括:

图像确定单元,用于依据目标链表的节点中的图像信息,确定所述目标链表中每个节点内的第二哈希值所属的候选图像。

11. 根据权利要求7或8所述的图像检索装置,其特征在于,所述哈希检索单元,包括:

哈希发送子单元,用于将所述预设数量个连续的第一哈希值段发送给多台从服务器,以便所述从服务器依据所述第一哈希值段在所述第一哈希值中的目标位置顺序,从所述从服务器负责检索的区域段内查询目标哈希表,所述区域段为数据库的存储空间中的部分存储区域;

哈希检索子单元,用于通过所述从服务器从所述目标哈希表的多个哈希段中,检索出与所述第一哈希值段相同的第二哈希值段;

所述图像匹配单元,包括:

结果接收单元,用于接收所述从服务器返回的查询结果,所述查询结果至少包括:从所述第二哈希值段所属的候选图像中,检索出的与所述参考图像相似的目标图像的信息。

12. 根据权利要求11所述的图像检索装置,其特征在于,还包括:

故障监测单元,用于在监测到所述多台从服务器中存在运行异常的故障从服务器的情况下,确定所述故障从服务器当前负责检索的区域段;

重分配单元,用于将所述故障从服务器对应的区域段分配给至少一个未出现运行异常的从服务器。

13. 一种服务器,其特征在于,包括:

处理器和存储器;

其中,所述处理器用于执行所述存储器中存储的程序;

所述存储器用于存储程序,所述程序至少用于:

获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈

希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像;

所述目标哈希表中存储有至少两条链表,所述链表中的节点用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段,且属于同一个链表的节点中存储的哈希值段相同;所述目标哈希表中的链表通过如下方式生成:

获取数据库中待处理的图像的哈希值;

将所述图像的哈希值拆分为所述预设数量个哈希值段,其中,不同的所述哈希值段在所述图像的哈希值中的位置顺序不同;

对于拆分出的每个哈希值段,定位与所述哈希值段的位置顺序对应的目标哈希表,并检测所述目标哈希表中是否存在用于记录所述哈希值段的链表;

在所述目标哈希表中存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成的节点添加到所述用于记录所述哈希值段的链表的尾端;

在所述目标哈希表中不存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成节点设定为所述目标哈希表中新增的链表的首个节点。

14. 一种可读存储介质,其特征在于,所述存储介质中存储有计算机程序,所述计算机程序被处理器加载并执行时,实现如权利要求1至6任一项所述的图像检索方法。

15. 一种图像检索系统,其特征在于,包括:

至少一台主控服务器,以及多台从服务器;

其中,所述主控服务器,用于获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;将所述预设数量个连续的第一哈希值段发送给所述多台从服务器;接收所述从服务器返回的查询结果,所述查询结果至少包括:所述从服务器检索出的与所述参考图像相似的目标图像的信息;

所述从服务器,用于依据所述第一哈希值段的位置顺序,从所述从服务器负责检索的区域段内定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述区域段为数据库的存储空间中的部分存储区域,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;从所述第二哈希值段所属的候选图像中,检索出与所述参考图像相似的目标图像;将检索出的目标图像的信息发送给所述主控服务器;

所述目标哈希表中存储有至少两条链表,所述链表中的节点用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段,且属于同一个链表的节点中存储的哈希值段相同;所述目标哈希表中的链表通过如下方式生成:

获取数据库中待处理的图像的哈希值;

将所述图像的哈希值拆分为所述预设数量个哈希值段,其中,不同的所述哈希值段在所述图像的哈希值中的位置顺序不同;

对于拆分出的每个哈希值段,定位与所述哈希值段的位置顺序对应的目标哈希表,并检测所述目标哈希表中是否存在用于记录所述哈希值段的链表;

在所述目标哈希表中存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈

希值段的节点,并将生成的节点添加到所述用于记录所述哈希值段的链表的尾端;

在所述目标哈希表中不存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成节点设定为所述目标哈希表中新增的链表的首个节点。

## 图像检索方法、装置、系统、服务器以及存储介质

### 技术领域

[0001] 本申请涉及图像检索技术领域,尤其涉及一种图像检索方法、装置、系统、服务器以及存储介质。

### 背景技术

[0002] 图像检索技术已经普遍应用于多种领域。如,对于购物平台而言,可能需要根据用户提供的待检索图像,从数据库存储的多种产品图像中,查找出与该待检索图像相似的产品图像,以定位到该产品图像对应的产品。又如,对于针对图像进行版权鉴定的版权鉴定平台,需要根据用户提供的待检索图像,查询数据库中是否存在与该待检索图像相似的图像,以为版权鉴定提供依据。

[0003] 在图像检索的过程中,为了查询数据库中是否存在与待检索图像相似的图像,需要依次计算该数据库中每幅图像与该待检索图像的相似性。而计算两幅图像之间相似性的复杂度较高,耗时相对较长;而且,由于数据库中图像的数量较大,依次计算数据库中每幅图像与该待检索图像的相似性,也必然导致数据计算量过大,从而使得图像检索效率过低。

### 发明内容

[0004] 有鉴于此,本申请提供了一种图像检索方法、装置、系统、服务器以及存储介质,以减少图像检索过程中的数据处理量,提高图像检索效率。

[0005] 为实现上述目的,一方面,本申请提供了一种图像检索方法,包括:

[0006] 获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

[0007] 将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

[0008] 根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

[0009] 从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像。

[0010] 另一方面,本申请还提供了一种图像检索装置,包括:

[0011] 哈希获取单元,用于获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

[0012] 哈希拆分单元,用于将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

[0013] 哈希检索单元,根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

[0014] 图像匹配单元,用于从所述第二哈希值段所属的候选图像中,检索与所述参考图

像相似的目标图像。

[0015] 另一方面,本申请还提供了一种服务器,包括:

[0016] 处理器和存储器;

[0017] 其中,所述处理器用于执行所述存储器中存储的程序;

[0018] 所述存储器用于存储程序,所述程序至少用于:

[0019] 获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

[0020] 将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

[0021] 根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

[0022] 从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像。

[0023] 另一方面,本申请还提供了一种可读存储介质,所述存储介质中存储有计算机程序,所述计算机程序被处理器加载并执行时,实现如前述的图像检索方法。

[0024] 另一方面,本申请还提供了一种图像检索系统,包括:

[0025] 至少一台主控服务器,以及多台从服务器;

[0026] 其中,所述主控服务器,用于获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;将所述预设数量个连续的第一哈希值段发送给所述多台从服务器;接收所述从服务器返回的查询结果,所述查询结果至少包括:所述从服务器检索出的与所述参考图像相似的目标图像的信息;

[0027] 所述从服务器,用于依据所述第一哈希值段的位置顺序,从所述从服务器负责检索的区域段内定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述区域段为数据库的存储空间中的部分存储区域,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;从所述第二哈希值段所属的候选图像中,检索出与所述参考图像相似的目标图像;将检索出的目标图像的信息发送给所述主控服务器。

[0028] 可见,在本申请实施例中,在获取到待检索的参考图像的哈希值之后,会先将该参考图像的哈希值拆分成预设数量个连续的哈希值段,并分别根据每个哈希值段在该参考图像的哈希值中的位置顺序,查询出该位置顺序上具有相同哈希值段的候选图像。而在两幅图像的海明距离小于预设阈值则确定两幅图像相似的前提下,该预设阈值实际上反映了这两幅图像的哈希值中相同位上具有不同数值的位数,因此,如果两幅图像有可能相似,将这两幅图像的哈希值拆分成多于该预设阈值个哈希值段,则这两幅图像至少会有一个哈希值段是完全相同,基于此原理可知,通过本申请的方案提取出的有可能与该参考图像相似的候选图像,相对于数据库中的所有图像,筛选出的候选图像的数据量很小,这样,只需计算参考图像与较少的候选图像之间的相似度,从而大大降低了图像检索的计算量。

## 附图说明

[0029] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例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申请的实施例,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,还可以根据提供的附图获得其他的附图。

[0030] 图1示出了本申请实施例中一种图像检索方法一种应用场景的组成结构示意图;

[0031] 图2a示出了本申请实施例中构建哈希表的一种流程示意图;

[0032] 图2b示出了将图像的哈希值拆分出的哈希值段分别存储到不同哈希表的示意图;

[0033] 图3示出了本申请一种图像检索方法一个实施例的流程示意图;

[0034] 图4示出了本申请实施例的图像检索方法所适用的又一种应用场景的组成结构示意图;

[0035] 图5示出了本申请实施例中主控服务器配置从服务器的一种流程示意图;

[0036] 图6示出了本申请实施例的一种图像检索方法又一种流程示意图;

[0037] 图7示出了本申请实施例的一种图像检索装置一种结构示意图;

[0038] 图8示出了本申请一种服务器的组成架构示意图。

## 实施方式

[0039] 本申请实施例的图像检索方法可以应用于多种不同的场景中,例如,应用于图片的版权鉴定,或者应用于产品图像定位等等场景。

[0040] 如,参见图1,其示出了本申请图像检索方法所适用的一种场景的组成结构示意图。

[0041] 在图1的场景中包括:图像检索系统100和终端200。该图像检索系统包括:应用服务器101,检索服务器102和数据库103。

[0042] 其中,终端200,用于向图像检索系统中的应用服务器101发送图像检索指令,该图像检索指令携带有待检索的图像;

[0043] 相应的,该应用服务器101,用于响应于该图像检索指令,并将该待检索图像发送给检索服务器102,并在检索服务器102返回图像检索结果时,将图像检索结果发送给该终端。如,该应用服务器101可以为web服务器。

[0044] 该检索服务器102用于根据依据待检索图像,检索数据库103中是否存储有与该待检索图像相似的图像,并将检索结果返回给该应用服务器。

[0045] 其中,数据库103中存储有多幅图像。

[0046] 其中,在本申请实施例中,图像可以有多种形式,如静态的、动态的,在一种可选的方式中,一幅图像可以认为是一张图片。

[0047] 需要说明的是,在图1中是以应用服务器和检索服务器为两个独立的服务器为例进行说明,但是在实际应用中,在提供图像检索服务的网络平台的数据处理量不大的情况下,该应用服务器和检索服务器也可以为一台服务器。

[0048] 可以理解的是,检索服务器分析两幅图像是否相似的方式可以有多种,在本申请实施例中,以检索服务器通过确定两幅图像的哈希值的海明距离,来分析两幅图像的相似性为例进行介绍。

[0049] 其中,图像的哈希值也就是俗称的图像的指纹,指依据图像所包含的特征,生成的一组哈希值。可见,图像的哈希值用于表征图像具有的特征。图像的哈希值可以为指定位数的二进制编码,该指定位数可以根据需要设定。如,提取图像的特征,生成一个64位的哈希值作为图像的指纹。在本申请实施例中,图像的哈希值是判断图像是否相似的依据。

[0050] 其中,本申请中确定表征图像的哈希值的方式可以有多种,如,可以采用感知哈希算法获得图像的哈希值,对此本申请实施例不加以限制。

[0051] 0海明距离也称为码距,在信息编码中,是指两个代码对应位置上不同编码的个数。换句话说,海明距离就是指将一串代码变换为另一串代码所需替换的字符个数。

[0052] 由于图像的哈希值实际上就是一个多位二进制编码,因此,在本申请实施例中,通过比较两幅图像的哈希值(或者说指纹)的海明距离,来判断两幅图像是否相似。具体的,可以设定两幅图像之间的海明距离小于预设阈值时,认为两幅图像相似。如,该预设阈值可以为3。

[0053] 其中,为了便于描述,在本申请实施例中将两幅图像的哈希值之间的海明距离也简称为两幅图像之间的海明距离。

[0054] 可以理解的是,为了基于海明距离,判断两幅图像之间的相似性,在本申请实施例中,数据库中除了存储多幅图像之外,还需要存储多幅图像各自对应的哈希值。相应的,应用服务器或者检索服务器可以计算出用于表征该待检索图像的特征的哈希值。而该检索服务器可以依据该待检索图像的哈希值,以及数据库中各幅图像的哈希值,依次计算该待检索图像与数据库中每幅图像之间的海明距离,以最终检索出与该待检索图像相似的图像。

[0055] 然而本申请的发明人发现:计算两幅图像之间的海明距离的计算复杂度相对较高,耗时相对较长;而且,由于数据库中图像的数量较多,如果依次计算该待检索图像与数据库中每幅图像之间的海明距离,必然导致图像检索的耗时较长,影响到图像检索效率。可见,如果减少计算海明距离的次数,则可以大大缩短计算耗时,提高图像检索效率。

[0056] 同时,发明人进一步研究发现:在两幅图像的哈希值之间的海明距离小于预设阈值,则认为这两幅图像相似的情况下,如果两幅图像相似,那么这两幅图像的N位哈希值中最多有该预设阈值R位不同。其中,N代表哈希值对应的二进制编码的位数,R表示预设阈值。如,假设预设阈值R为3,假设图像的哈希值的位数N为64位,那么两幅图像的哈希值中最多会有3位不同,例如,有可能这两个哈希值上的第3、20和63位上不同。

[0057] 由此可知,如果两幅图像相似,在分别将每幅图像的N位哈希值拆分为R+1段,得到每幅图像对应的R+1个哈希值段之后,这两幅图像各自对应的R+1个哈希值段中,至少有一个相同位置上的哈希值段是完全相同的。举例说明,假设R为3,R+1为4,那么图像1和图像2相似的话,图像1的哈希值X和图像2的哈希值Y中最多有3位上的数值不同,如果将哈希值X和哈希值Y分别拆分为4个(即R+1段)哈希值段,那么哈希值X和哈希值Y中最多3位不同的数值也最多只能分布到3个哈希值段中,因此,哈希值X的4个哈希值段与哈希值Y的4个哈希值段中至少有一个相同位置上的哈希值段是相同的。

[0058] 经由以上分析可知,如果两个图像各自对应的R+1个哈希值段均不相同,那么这两个图像肯定不相同;而如果两个图像的各自对应的R+1个哈希值段存在至少一个相同位置上的哈希值段相同,则说明这两幅图像有可能相似。

[0059] 需要说明的是,以上是以将图像的哈希值拆分为R+1段为例进行说明,但是可以理

解的是,在基于两幅图像的哈希值小于预设阈值R的情况下,如果两幅图像相似,将两幅图像的哈希值拆分为预设数量个哈希值段,只要该预设数量大于该预设阈值,这两幅图像的哈希值拆分为的哈希值段中,同样至少有一个相同位置上的哈希值段是相同的。

[0060] 因此,本申请实施例在计算待检索图像的哈希值与数据库中不同图像的哈希值之间的海明距离之前,会先将待检索图像的N位哈希值拆分为预设数量个哈希值段,并依据该待检索图像的预设数量段哈希值段,以及数据库中多个图像对应的各自对应的预设数量个哈希值段,查询出数据库中与该待检索图像具有至少一个相同的哈希值段的图像,以排除数据库中不可能与待检索图像相似的图像。由于比较两个哈希值段相同所需的耗时较少,相对于计算海明距离的耗时较小,而数据库中与该待检索图像具有至少一个相同的哈希值段的图像的数量,远远小于该数据库中图像的总数量,这样,在计算查询出的图像与该待检索图像的海明距离,便可以大大减少计算海明距离的次数,从而有利于提高图像检索效率。

[0061] 基于以上研究发现,在本申请实施例中,数据库中不仅需要存储多幅图像以及每幅图像的N位哈希值之外,还需要分别存储每幅图像的哈希值拆分为的预设数量(如,R+1)个哈希值段。其中,N表示哈希值的位数,具体可以根据需要设定。其中,该预设数量为大于该预设阈值R的自然数就可以。为了便于描述以下以预设数量为R+1为例,该R+1为预设阈值R加1,该预设阈值R为预先设定的,用于表征两幅相似图像的海明距离所需小于的数值。

[0062] 为了便于理解,先介绍图像的哈希值拆分为R+1段哈希值段的过程。

[0063] 其中,将图像的N位哈希值拆分为R+1个哈希值段是按照该N位哈希值中各个位的前后顺序,将该N位哈希值的N位上的数值拆分到多个段中,并将拆分出的每段称为一个哈希值段。可以理解的是,每个哈希值段包括哈希值中的至少一位的数值。

[0064] 根据实际需要的不同,每幅图像的N位哈希值的R+1个哈希值段中包含的位数可以相同,也可以是不同的,只要保证按照同一种哈希值拆分方式拆分不同图像的哈希值,以使得不同图像中对应位置的哈希值段所包含的位数相同即可。

[0065] 举例说明,以哈希值为64位,R+1为4为例,则对于任意一个图像的哈希值,可以是将哈希值平均拆分为4个哈希值段,每个哈希值段包括16位数据,这样,不同图像的哈希值段也均包括16位数据;也可以将该哈希值的前8位作为第一个哈希值段,将该哈希值的第9到32位作为第二个哈希值段,将第33到40位作为第三个哈希值段,并将第41到64位作为第四个哈希值段,这样,对于任意两个图像而言,这两个图像的第一个哈希值段中包含的位数相同,相应的,这两个图像的第二、第三和第四个哈希值段中包含的位数也分别相同。

[0066] 可以理解的是,为了便于拆分,将N位哈希值平均拆分为R+1段哈希值段为一种优选实施方式。其中,如果N位能被R+1整除,那么每个哈希值段中包含的位数相同,如,哈希值为64位,R+1为4,则每个哈希值段包含16位数值。如果N位不能被R+1位整除,那么不同哈希值段中包含的位数最多差一个,如,哈希值为64位,R+1为5,则前4个哈希值段均包括13位数值,而最后一个哈希值段包含12个哈希值段。

[0067] 以上介绍了将图像的哈希值拆分为哈希值段的过程。可以理解的是,对于待存储到数据库中的图像而言,将该图像的N位哈希值拆分为R+1个哈希值段之后,可以将该图像、该图像的N位哈希值以及该图像对应的R+1个哈希值段关联存储,以为后续图像检索提供依据。如,在一种实现方式中,数据库中存储图像的同时,会将该图像的N位哈希值、R+1个哈希值段作为该图像的关联信息进行存储。

[0068] 考虑到数据库中图像的数量较多,如果依次将待检索图像对应的R+1个哈希值段分别与数据库中所有图像在相对位置(相同位置顺序)上的哈希值段进行比较,比较次数较多,也需要耗费一定的时长。为了进一步降低检索耗时,本申请实施例中除了向数据库中存储图像以及图像的哈希值之外,还需要根据图像所具有的哈希值段的数量,生成多个哈希表,每个哈希表用于记录不同图像在同一个位置上的哈希值段的信息。

[0069] 如,参见图2a,其示出了本申请实施例中生成存储哈希值段的哈希表的流程示意图,该流程可以由应用服务器或者检索服务器执行,该流程可以包括:

[0070] S201,构建预设数量R+1个指针数组,每个指针数组作为一个哈希表。

[0071] 其中,R+1表示预设数量,该预设数量为前面提到的针对海明距离设置的预设阈值R加1。

[0072] 其中,指针数组的长度可以根据需要设定,如,可以采用 $2^{16}$ 长度的指针数组。

[0073] 在本实施例中,每个指针数组作为一个哈希表用于存储图像的哈希值拆分出的某个位置顺序上的哈希值段。如,假设每个图像的哈希值拆分为4个哈希值段,则说明图像的哈希值具有4个顺序相连的哈希值段,这4个哈希值段的位置顺序不同,且位置顺序依次为哈希值的第一个哈希值段、第二个哈希值段、第三个哈希值段和第四个哈希值段;相应的,需要构建4个哈希表,这4个哈希表分别用于记录第一个、第二个、第三个以及第三个哈希值段。

[0074] S202,获取数据库中待处理的图像的哈希值,并将该图像的哈希值拆分为R+1个哈希值段。

[0075] 其中,每个哈希值段对应该图像的哈希值中的一个位置顺序,不同哈希值段在该哈希值中的位置顺序不同。哈希值段的位置顺序表征该哈希值段在所属的哈希值中的位置顺序,也是该哈希值段在该多个连续的哈希值段中顺序,如,哈希值段为该哈希值中从前到后的第二个哈希值段,该哈希值段的位置顺序可以是第二个。

[0076] 其中,图像的哈希值可以通过提取该图像的特征得到,本申请对于如何提取图像的哈希值不做限定。将图像的哈希值拆分为哈希值段的过程可以参见前面的相关介绍,在此不再赘述。

[0077] 在该步骤S202可以是:在数据库已经存储了图像以及图像的哈希值的基础上,获取数据库已经存储的图像的哈希值,以便执行对该哈希值的拆分,进而生成哈希表;也可以是:在搜索到待存储到数据库中的图像,先向数据库中存储该图像以及图像的哈希值,然后获取当前存储到数据库中的图像的哈希值,并将该图像的哈希值进行拆分。

[0078] S203,对于任意一个哈希值段,根据该哈希值段在哈希值中的位置顺序,确定该位置顺序对应的哈希表,并检索该哈希表中是否存在用于记录该哈希值段的链表,如果是,则执行步骤S204;如果否,则执行步骤S205。

[0079] 位置顺序对应的哈希表用于存储不同图像的哈希值中处于该位置顺序的哈希值段的信息。如,根据哈希值段的位置顺序,确定出该哈希值段为图像的哈希值中第二个哈希值段,则需要确定用于存储第二个哈希值段的哈希表。

[0080] 其中,每个哈希表所表征的指针数组中均存储有至少一个链表,每个链表包括多个节点(也称为结点),每个节点中记录一个哈希值段,还可以记录有该哈希值段所属的图像信息,如,哈希值段所属的图像的标识,或者哈希值所属的图像的哈希值。其中,同一个哈

希表中不同节点内存储的哈希值段属于不同的图像。当然,该节点中还具有指向该链表中下一个节点的指针。

[0081] 在本申请实施例中,一个链表中各个节点分别用于存储不同图像的哈希值对应的同一个位置顺序上的哈希值段,且同一个链表中各个节点内的哈希值段相同。这样,由于链表中各个节点内的哈希值段,后续依据待搜索的哈希值段,从该哈希表中搜索相同哈希值段时,则可以直接定位包含该哈希值段的链表,根据链表中各个节点包含的图像信息,便可以找出相应位置顺序上与该哈希值段相同的所有图像。

[0082] 相应的,在确定出该位置顺序对应的哈希表之后,需要检测该哈希表中是否存在用于记录该哈希值段的链表,以便将图像的该哈希值段记录到该链表中节点中,以实现将不同图像的哈希值对应位置顺序的哈希值段均归类到一起。

[0083] 可以理解的是,为了便于区分,可以为每个链表构建一个索引,索引中用于表征该链表中各个节点内记录的哈希值段的内容,如,该索引还可以为该链表中节点内存储的哈希值段。当然,也可以直接将链表中第一节内哈希值段作为链表的索引。

[0084] S204,生成用于表征该哈希值段的节点,并将该节点添加到该链表的尾端。

[0085] 为了实现归类,如果存在用于记录该哈希值段的该链表,则可以生成包含该哈希值段的节点,并将该节点添加到该链表的尾端。

[0086] S205,生成用于表征该哈希值段的节点,并将该节点作为该哈希表中新增设的链表的首个节点。

[0087] 相应的,如果不存在用于记录该哈希值段的链表,则说明当前时刻哈希值表中还未存储过相同的该哈希值段,在该种情况下,可以创建一个用于记录该哈希值段的链表,即将生成的包含该哈希值段的节点作为一个链表的首个节点。

[0088] 通过以上步骤S202到步骤S205可以实现将图像的哈希值段的信息存储到哈希表中,从而得到分别用于存储图像的哈希值对应的不同位置顺序的哈希表。

[0089] 为了便于理解图2a的流程,举例说明,假设图像的哈希值为64位,并假设将图像的哈希值拆分为4段,得到四个哈希值段,参见图2b,该图2b示出了将图像的哈希值拆分出的四段哈希值段存储到哈希表中的一个示例。

[0090] 由图2b可知,假设图像的64位哈希值为:110 1000 0111 0011 1001 0110 1110 1010 1111 0001 1100 1001 1110 0001 0001 1111。则该图像的哈希值拆分出的第一个哈希值段为:0110 1000 0111 0011;第二个哈希值段为1001 0110 1110 1010;第三个哈希值段为1111 0001 1100 1001;第四个哈希值段为1110 0001 0001 1111。

[0091] 相应的,需要将该图像的第一个哈希值段存储到对应用于存储前1-16位构成的哈希值段的哈希表中,如图2b中,第一个哈希表,即标示有前1-16位的哈希表。

[0092] 进一步的,还需要检测该哈希表中是否存在用于存储“0110 1000 0111 0011”的链表。在图2b将链表中存储的哈希值段的具体取值作为该链表的索引,如参见图2b的第一哈希表中各行行为一条链表,这样,可以直接定位索引为“0110 1000 0111 0011”的链表,如图2b中该图像的第一个哈希段通过箭头指向的链表。相应的,为该图像生成用于存储该第一哈希值段的节点,并在节点中存储该图像的相关信息,如该图像的哈希值以及在数据库中存储位置等,并将该节点添加到定位出的链表中即可。当然,如果未定位出到索引为“0110 1000 0111 0011”的链表,则可以将该节点独立作为一个链表,在此不再赘述。

[0093] 可以理解的是,在图2b中,对于第2-4个哈希表中并未明确画出哈希表内的链表,但是,将该图像的第二个、第三个以及第四个哈希值段可以分别存储到第二个、第二个和第四个哈希表中,且具体过程与存储第一个哈希值段的过程相似,在此不再赘述。

[0094] 可以理解的是,在图2a的流程中,是以任意一个位置顺序对应的哈希表中,将该位置顺序上的哈希值段不同的不同图像的信息存储到一个链表为例说明。但是可以理解的是,如果仅仅构建不同位置顺序对应的哈希表,并将图像的哈希值对应的不同位置顺序上哈希值段以及相关信息分别存储到各自对应的哈希表中也同样适用于本实施例,如,在该步骤S203中,确定出确定该位置顺序对应的哈希表之后,可以将该哈希值段以及该哈希值段所属的图像存储到该位置顺序对应的该哈希表中。

[0095] 需要说明的是,本实施例是以每个哈希表为一个指针数组为例进行介绍,但是可以理解的是,将通过其他数据存储形式来存储哈希表也同样适用于本申请实施例;如,每个哈希表也可以认为一个文件夹,该文件夹中存储有后续的一个或多个链表。

[0096] 结合以上共性,对本申请实施例的图像检索方法进行介绍,参见图3,其示出了本申请一种图像检索方法一个实施例的流程示意图,本实施例的方法可以包括:

[0097] S301,应用服务器获取待检索的参考图像,并计算该参考图像的哈希值。

[0098] 如,应用服务器可以接收客户端发送的检索指令,该检索指令携带有该待检索的图像。又如,应用服务器获取指定存储空间内预置的待检索的图像,当然,对于不同的场景,该应用服务器获取该待检索图像的方式可以有所不同。

[0099] 在本申请实施例中,为了便于区分,将待检索的图像称为参考图像。参考图像的哈希值为表征该参考图像的图像特征的二进制码。

[0100] S302,应用服务器将该参考图像的哈希值发送给检索服务器。

[0101] 如,应用服务器向该检索服务器发送检索请求,在该检索请求中携带该参考图像的哈希值。

[0102] S303,检索服务器将该参考图像的哈希值拆分成预设数量个连续的哈希值段。

[0103] 其中,每个哈希值段包括至少一位,且该预设数量个连续的哈希值段可以组合成该参考图像的哈希值。

[0104] 其中,将参考图像的哈希值拆分成连续的预设数量个哈希值段的过程可以参见前面拆分图像的哈希值的过程,在此不再赘述。

[0105] 需要说明的是,图3的实施例中,是以应用服务器计算出待检索的参考图像的哈希值为例进行说明,但是可以理解的是,应用服务器也可以直接将该参考图像发送给该检索服务器,并由该检索服务器确定该参考图像的哈希值。

[0106] S304,对于参考图像的哈希值拆分出的每个哈希值段,检索服务器依据该哈希值段在该参考图像的哈希值中的目标位置顺序,定位数据库中与该目标位置顺序对应的目标哈希表。

[0107] 在本申请实施例中,为了便于区分,将当前处理的哈希值段的位置顺序称为目标位置顺序。

[0108] 可以理解的是,该步骤S304可以从数据库中查询该目标哈希表;在检索服务器缓存了哈希表的前提下,也可以是从终端缓存的哈希表中定位出该目标哈希表。

[0109] S305,检索服务器检测该目标哈希表中,是否存在用于存储该哈希值段的目标链

表,如果是,则执行步骤S306;如果否,则执行步骤S307。

[0110] 其中,考虑到数据库中图像的复杂度,一般情况下,每个哈希表中会包括至少两条链表。

[0111] 其中,目标链表中各个节点中存储的哈希值段均与参考图像对应的哈希值段相同。

[0112] S306,检索服务器分别确定该目标链表中各个节点内的哈希值段所属的候选图像以及该候选图像的哈希值。

[0113] 可以理解的是,如果检测到目标链表,则说明数据库中存在该目标位置顺序上的哈希值段与参考图像中该目标位置顺序上的哈希值段相同的图像。为了便于区分,将目标链表中各个节点内的哈希值段所对应的图像称为候选图像。可以理解的是,候选图像与参考图像至少一个位置顺序上的哈希值段相同,因此,候选图像有可能是与参考图像相似的图像。

[0114] 其中,确定候选图像的哈希值可以从数据库中获取确定出的候选图像的哈希值,如果该节点中存储有该候选图像的哈希值,则可以直接获取该候选图像的哈希值。

[0115] S307,检索服务器检测该参考图像对应的多个哈希值段中,是否存在未被检索的哈希值段,如果是,则返回该步骤S304,以依据未被检索的哈希值段的目标位置顺序,确定目标哈希表。

[0116] 可以理解的是,对于参考图像对应的某个位置顺序上的哈希值段而言,如果从相应的目标哈希表中未检索到相应的目标链表,则说明数据库中不存在该位置顺序上具有该哈希值段的其他图像。在该种情况下,则可以检测是否该参考图像对应的其他哈希值段均被检索,以便完成所有参考图像所有哈希值段的检索。

[0117] 可以理解的是,对于每个哈希值段均需要执行以上步骤S304、S305和S306,且可以依次是检索每个哈希值段,也可以并行执行各个哈希值段的检索。

[0118] 可以理解的是,在实际应用中,也可以在S305和S306之后不执行步骤S307,而直接是在检索到候选图像之后,直接执行后续步骤S308。如,在步骤S305中未检索到该哈希值段对应的目标链表,则可以结束该哈希值段的处理,但仍执行依据其他哈希值段检测到目标链表的处理;又如,对于采用并行方式检索参考图像的各个哈希值段的情况中,各个哈希值段均同时被检索,所以可以无需执行该步骤S307。

[0119] 需要说明的是,对于直接在某个位置顺序对应的哈希表中直接存储不同图像在该位置顺序上的哈希值段,也同样适用于本实施例。相应的,在步骤S304确定出目标哈希表之后,可以直接从该目标哈希表中检索是否存在与该参考图像对应的该哈希值段相同的候选哈希值段,以便后续确定候选哈希值段所属的候选图像以及候选图像的哈希值等。

[0120] 需要说明的是,在本申请实施例中,为了便于区分参考图像与候选图像的哈希值,也可以将参考图像的哈希值称为第一哈希值,并将第一哈希值拆分出的哈希值段称为第一哈希值段;而将从哈希表中定位出与该第一哈希值段相同的哈希值段称为第二哈希值段,即,候选图像的哈希值称为第二哈希值,将第二哈希值拆分出的哈希值段称为第二哈希值段。

[0121] S308,检索服务器依据参考图像的哈希值以及候选图像的哈希值,分别计算参考图像与每个候选图像的海明距离。

[0122] 可以理解的是,两幅图像的哈希值中相同位置上不同字符的个数就是海明距离,而本申请中计算海明距离的方式可以有多种,在此不加以限制。

[0123] 可以理解的是,依据参考图像的一个哈希值段有可能会定位到多个候选图像,因此,该步骤S308需要分别计算参考图像与定位出的每个候选图像的海明距离。

[0124] S309,检索服务器将与参考图像的海明距离小于预设阈值的候选图像,确定为与该参考图像相似的图像。

[0125] 可以理解的是,前面步骤S303到S307为依据海明距离确定两幅图像相似性的原理,先从数据库中确定出与参考图像有可能相似的候选图像。在确定出候选图像之后,可以按照步骤S308和S309的方式,以检索出与参考图像相似的目标图像;也可以是通过其他方式从确定出的候选图像中,检索出与该参考图像相似的目标图像,对此不加以限制。

[0126] 需要说明的是,由于依据参考图像的不同哈希值段,有可能会同时定位到一个候选图像,如,如果参考图像与候选图像的哈希值完成相同,使得依据参考图像的哈希值拆分出的各个哈希值段,均能定位到该参考图像,为了避免重复操作,在确定出所有候选图像之后,还可以先去除重复的候选图像。或者是,如果计算过该候选图像与参考图像的海明距离,后续在定位到该候选图像时,则无需重复计算该海明距离。

[0127] S310,检索服务器依据与该参考图像相似的图像,向应用服务器反馈检索结果。

[0128] 如,将与该参考图像相似的图像发送给应用服务器;或者,将与该参考图像相似的图像的个数(还可以有图像来源等相关信息)发送给应用服务器等。

[0129] 可见,在两幅图像相似的前提为这两幅图像的海明距离小于预设阈值的前提下,如果将这两幅图像的哈希值拆分成预设阈值加1个(或者是加任意不小于的自然数)哈希值段,则这两幅图像至少会有一个哈希值段是完全相同,基于此原理,本申请实施例在获取到待检索的参考图像的哈希值之后,会先将该参考图像的哈希值拆分成预设数量个连续的哈希值段,并根据哈希值段在该参考图像的哈希值中的位置顺序,并检索该位置顺序上具有相同哈希值段的候选图像,从而可以从数据库中筛选出有可能与该参考图像相似的候选图像,并剔除了完全不可能与该参考图像相似的其他图像,而且,由于检索位置顺序上具有哈希值段的图像的数据处理量较小,相对于计算相似性的计算量几乎可以忽略。而相对于数据库中的所有图像,筛选出的候选图像的数据量又很小,这样,只需计算参考图像与较少的候选图像之间的相似度,从而大大降低了图像检索的计算量。

[0130] 同时,由于同一个哈希表存储的是不同图像的哈希值中同一个位置顺序上的哈希值段,且该将该位置顺序上具有相同哈希值段的不同图像的信息分别存储到一个链表的各个节点中,因此,根据参考图像的一个哈希值段,可以直接定位出该位置顺序上的哈希值段与参考图像中该哈希值段相同的所有节点所对应的候选图像,从而进一步减少了筛选出候选图像的耗时。

[0131] 为了便于理解本申请的有益效果,举例说明:

[0132] 假设数据库中有1亿幅图像的哈希值,如果依次计算参考图像与数据库中各幅图像的海明距离,需要计算1亿次海明距离。

[0133] 而假设每个哈希值为64位,且海明距离小于3才可以认为两幅图像相似,采用本申请的方案,每幅图像的哈希值被拆分为4个16位的哈希值段,并分别存储到4个哈希表中,且由于每个哈希表中采用链表形式存储属于不同图像却相同的哈希值段,因此,在将参考图

像的哈希值拆分为4个哈希值段之后,对于每个16位的哈希值段,筛选候选图像时仅仅是哈希查找,便可以定位到链表中各个节点所表征的候选图像,耗时的复杂度为 $O(1)$ 。而图像检索的耗时主要取决于:计算参考图像与链表中各个节点表征的候选图像之间海明距离的次数。而以哈希表中每个链表为 $2^{16}$ 的指针数组为例, $2^{16}$ 为65536位,由于每个哈希表需要存储该1亿幅图像的哈希值中的一段哈希值段,那么每个链表的长度为1亿/65536=1526次,而由于参考图像对应4个哈希值段,则最多查询到4个链表,因此,海明距离的最大计算次数为 $1526*4=6104$ 次,而6104次远远小于1亿次,检索效率大大提高。

[0134] 在本申请以上实施例中,是以检索服务器为单台服务器为例介绍,但是为了适应数据处理较大的图像检索场景,提高图像检索效率,在实际应用中,还可以将该检索服务器替换为多台检索服务器构成的分布式系统。对于多台检索服务器构成的服务器集群的情况,其与单台检索服务器的操作相似,只不过每台检索服务器所负责检索的图像会有所差异,但是对于一幅图像的检索过程是相同的。

[0135] 可以理解的是,由于每幅图像的哈希值所拆分出的多个哈希值段会被存储到多个哈希表中,且每个哈希表中还需要存储该图像的信息,而使得检索服务器需要加载到内存中数据量较大。如,以节点存储哈希值段以及哈希值段所属的图像的相关信息为例,每个节点可能需要占用64比特的空间,而1亿幅图像就需要6.4G的空间。如果采用单台检索服务器,则可能导致该检索服务器内存不足的情况,且一旦出现故障,将会导致无法提供检索服务;而采用多台检索服务器构成的分布式集群,则可以降低内存不足,并提高检索服务的稳定性。

[0136] 为了能够更为合理的为服务器集群中各台检索服务器分配检索任务,在本申请实施例中,检索服务器集群中可以包括至少一台主控服务器以及多台从服务器。如,参见图4,其示出了本申请图像检索方法所适用的又一种场景的组成结构示意图。

[0137] 本申请实施例的场景中:图像检索系统410和终端420。该图像检索系统包括:应用服务器411,检索服务器集群412和数据库413。

[0138] 其中,检索服务器集群412中包括主控服务器4121和多台从服务器4122。

[0139] 其中,主控服务器4121可以将检索任务分配给多台从服务器,以通过多台从服务器完成检索任务对应的检索。

[0140] 其中,主控服务器4121为从服务器分配检索任务的方式可以有多种。

[0141] 在一种可能的情况中,主控服务器可以配置各个从服务器负责检索的数据库的存储区域,以将数据库的各个存储区域从逻辑上划分给该多个从服务器。相应的,从服务器执行检索任务时,只从该从服务器被分配的数据库的存储区域内进行检索。为了便于理解,先介绍主控服务器配置从服务器的配置流程,如,参见图5,该配置流程可以包括:

[0142] S501,主控服务器确定当前处于正常运行的多个从服务器。

[0143] 如,主控服务器可以基于心跳机制,检测各个从服务器是否正常运行(俗称的处于存活状态)。例如,主控服务器可以定时向所有从服务器发送心跳消息,如果从服务器接收到该心跳消息,则从服务器会向主控服务器反馈一个心跳响应消息,这样,主控服务器接收到从服务器的心跳响应消息,则可以确定从服务器处于存活状态。

[0144] 相应的,主控服务器可以存储用于记录各个从服务器的存活状态的文件,以便依据该文件确定当前存活的从服务器。

[0145] S502,主控服务器将数据库的存储空间从逻辑上划分为多个区域段,并将该多个区域段分配给多个从服务器,以使得处于正常运行的每个从服务器至少负责一个区域段。

[0146] 其中,该数据库的存储空间可以为全部的存储空间,也可以为数据库中存储有数据的存储空间。在本申请实施例中,将数据库的存储空间划分为多个部分小存储空间,每个小存储空间称为一个区域段。其中,各个区域段对应的存储空间大小可以相同,也可以不同。

[0147] 如,主控服务器根据该处于正常运行状态的多个从服务器的总数量,将该数据库的存储空间划分为该总数量个区域段,并分别分配给每个从服务器,这样,每个从服务器负责数据库中的一个区域段。

[0148] 又如,主控服务器可以将该多个从服务器分为多组服务器组合,每组服务器组合包括至少两个从服务器,这样,主控服务器可以根据每组服务器组合中从服务器的数量,将数据库的存储空间划分为该数量个区域段,并将划分出的区域段分别分配该服务器组内的各个服务器。

[0149] 当然,还可以有其他为从服务器配置区域段的可能方式,在此不加以限制。

[0150] 可以理解的是,该主控服务器中还可以存储各个从服务器的配置文件,配置文件中还可以记录从服务器的内存空间、性能等参数,这样,在为从服务器分配数据库的区域段时,可以结合从服务器自身的参数,确定分配给从服务器的区域段对应的存储空间的大小。

[0151] S503,主控服务器将从服务器负责的区域段的地址信息,发送给从服务器。

[0152] 如,主控服务器将数据库的存储空间从地址0-10的区域段分配给一个从服务器,而将地址11-20的区域段分配给另一个从服务器,依次类推。相应的,对于分配地址0-10的区域段的从服务器而言,主控服务器可以将分配给从服务器对应的该区域段对应的地址0-10的信息发送给该从服务器。

[0153] S504,从服务器根据分配的区域段的地址信息,从数据库中读取该从服务器负责的区域段内的哈希表,并向从服务器的内存中加载获取到的哈希表。

[0154] 可以理解的是,在从服务器获取到分配的区域段的地址信息之后,从服务器可以获取该区域段内存存的哈希表,并加载到内存中,与完成哈希表的初始化。这样,后续从服务器可以查询内存中缓存的哈希表,以减少访问数据库的次数。

[0155] 当然,在实际应用中,从服务器也可以不执行该步骤S504,则是当接收到主控服务器指示的待检索的哈希值段之后,再依据该哈希值段从数据库中该区域段内查询相应的哈希表。

[0156] S505,当从服务器完成哈希表加载时,向主控服务器发送初始化完成消息。

[0157] S506,主控服务器确定接收到该多个从服务器各自发送的初始化完成消息时,确定从服务器的配置完成。

[0158] 以上步骤S505和S5056是为了便于主控服务器了解到配置进度,以保证所有从服务器均完成配置。

[0159] 可以理解的是,在主控服务器完成从服务器的配置之后,为了提高检索系统的稳定性,主控服务器仍会监测从服务器的运行状态是否正常,如,主控服务器可以通过前面提到的心跳机制监测从服务器的运行状态。在主控服务器检测到该从服务器出现运行异常(如,宕机等),该主控服务器会先确定该从服务器负责的区域段;然后,该主控服务器会将

确定出的该区域段分配给该从服务器之外的至少一个从服务器,以提高检索系统的稳定性。

[0160] 如,将确定出的该区域段平均分配给该从服务器之外的其他未出现异常的从服务器,以使得其他服务器被分配的区域段的容量增加。而从服务器在获取到从服务器新分配的区域段之后,仍可以按照前面S504和S504的步骤执行相关操作。

[0161] 可以理解的是,在主控服务器完成从服务器的配置之后,主控服务器便可以接收应用服务器发送的待检索的参考图像或者参考图像的哈希值。如,参见图6,其示出了本申请又一种图像检索方法又一个实施例的流程示意图,本实施例的方法中以检索服务器实现图像检索为例说明,本实施例可以包括:

[0162] S601,应用服务器获取待检索的参考图像,并计算该参考图像的哈希值。

[0163] S602,应用服务器将该参考图像的哈希值发送给检索服务器集群中的主控服务器。

[0164] 与前面实施例相似,应用服务器也可以直接将参考图像发送给主控服务器,由主控服务器计算参考图像的哈希值。当然,在实际应用中,应用服务器和主控服务器也可以为同一台服务器。

[0165] S603,主控服务器将该参考图像的哈希值拆分成预设数量个连续的哈希值段。

[0166] 可以理解的,该参考图像的哈希值也可以称为第一哈希值,而该参考图像的哈希值拆分出的哈希值段也可以称为第一哈希值段。

[0167] 以上步骤S601到S603可以参见前面实施例的相关介绍,在此不再赘述。

[0168] S604,主控服务器将该参考图像对应的该预设数量个连续的哈希值段发送给每个从服务器。

[0169] 可以理解的是,由于每个从服务器分别负责数据库的存储空间对应的一个区域段,因此,将该预设数量个连续的哈希值段发送给该多个从服务器,可以通过该多个从服务器并行的从数据库的不同区域段内进行搜索,从而可以提高搜索效率。

[0170] S605,从服务器对于参考图像对应的每个哈希值段,依据该哈希值段在该参考图像的哈希值中的目标位置顺序,从数据库中属于该从服务器负责的区域段中,确定与该目标位置顺序对应的目标哈希表。

[0171] 可以理解的是,由于每个从服务器负责数据库的一个区域段,因此,每个从服务器仅仅从其负责的区域段进行检索,以定位出该区域段内存储的与该目标位置顺序对应的目标哈希表。该过程与前面定位目标哈希表的过程相似,只不过每个从服务器从数据库中检索的范围缩小。

[0172] 当然,在实际应用中,从服务器负责的区域段内有可能并不存在与某个哈希值段的目标位置顺序对应的目标哈希表,则说明该从服务器并不负责该目标哈希表,而是由其他从服务器负责该目标哈希表,在该种情况下,该从服务器可以忽略该哈希值段对应的候选图像的相关检索,并可以继续检索是否存在其他哈希值段对应的目标哈希值表即可。

[0173] 需要说明的是,在本实施例中是以从服务器实时从数据库中查询是否存在目标哈希表为例进行说明,在实际应用中,如果在主服务器为从服务器分配区域段时,该从服务器已经加载了该区域段内对应的哈希表,则该从服务器可以直接从内存中存储的(也就是属于该区域段内)的哈希表中,检索是否存在相应的目标哈希表。

[0174] S606,从服务器从该目标哈希表中,定位出用于存储该哈希值段的目标链表。

[0175] S607,从服务器分别确定该目标链表中各个节点内的哈希值段所属的候选图像以及该候选图像的哈希值。

[0176] 如,从服务器从节点中查询出哈希值段所属的候选图像的存储地址,然后,依据该存储地址从数据库中查询出该候选图像的哈希值。

[0177] 以上步骤S606和S607可以参见前面实施例中检索服务器执行相关操作的具体实现,在此不再赘述。

[0178] 其中,为了便于区分,目标链表的节点内的哈希值段也可以称为第二哈希值段。

[0179] 可以理解的是,以上是以哈希表中包括链表为例进行说明,在哈希表中直接存储各个哈希值段的信息以及候选图像的情况下,从服务器还可以直接从目标哈希表中,确定出与该参考图像对应的哈希值段相同的第二哈希值段以及该第二哈希值段所属的候选图像的信息,具体过程可以参见前面实施例的相关描述。

[0180] S608,从服务器依据参考图像的哈希值以及候选图像的哈希值,分别计算参考图像与每个候选图像的海明距离。

[0181] 其中,以上步骤S605到S608中从服务器的操作可以与前面通过独立的检索服务器或者服务器集群中独立的检索服务器的操作过程相似,在此不再赘述。

[0182] 当然,在实际应用中,由于从服务器数量相对较多,通过从服务器计算海明距离从服务器可以提高计算效率。但是从服务器也可以在定位出候选图像之后,将定位出的候选图像的哈希值发送给主控服务器,由主控服务器计算参考图像与各个候选图像的海明距离。

[0183] S609,从服务器从候选图像中,确定出与参考图像之间的海明距离小于预设阈值的目标图像,并将目标图像的信息反馈给主控服务器。

[0184] S610,主控服务器依据不同从服务器返回的目标图像的信息,向应用服务器反馈检索结果。

[0185] 如,主控服务器将各个从服务器返回的目标图像的信息均发送给应用服务器;或者是,将从服务器返回的目标图像的总个数发送给应用服务器;还可以是,将目标图像的来源、标识以及总个数发送给应用服务器,当然,还可以其他反馈检索结果的方式,本申请对此不加以限制。

[0186] 在本申请的另一方面,对应本申请的图像检索方法,本申请还提供了一种图像检索装置。

[0187] 如,参见图7,其示出了本申请一种图像检索装置一个实施例的组成结构示意图,本实施例的装置可以应用于前面提到的检索服务器或者检索服务器集群。该装置可以包括:

[0188] 哈希获取单元701,用于获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

[0189] 哈希拆分单元702,用于将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

[0190] 哈希检索单元703,根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述

目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

[0191] 图像匹配单元704,用于从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像。

[0192] 在一种可能的实现方式中,该图像匹配单元可以包括:

[0193] 哈希确定单元,用于获取所述第二哈希值段所属的候选图像的第二哈希值,所述第二哈希值用于表征候选图像的图像特征;

[0194] 距离计算单元,用于依据所述参考图像的第一哈希值和所述候选图像的第二哈希值,计算所述候选图像与参考图像的海明距离;

[0195] 距离匹配单元,用于从所述第二哈希值段所属的候选图像中,检索出与所述参考图像的海明距离小于预设阈值的目标图像。

[0196] 在一种可能的实现方式中,所述哈希检索单元定位出的所述目标哈希表中存储有至少两条链表,每条链表包括至少一个节点,所述节点用于存储图像的哈希值中处于所述目标位置顺序的哈希值段,且属于同一个链表的节点中存储的哈希值段相同;

[0197] 所述哈希检索单元,包括:

[0198] 链表定位子单元,用于从所述目标哈希表的至少两条链表中,定位出用于存储所述第一哈希值段的目标链表,所述目标链表的各个节点中存储的第二哈希值段与所述第一哈希值相同。

[0199] 进一步的,所述链表的节点中还可以存储有所述节点内的哈希值段所属的图像的信息。

[0200] 相应的,该装置还可以包括:图像确定单元,用于依据目标链表的节点中的图像信息,确定所述目标链表中每个节点内的第二哈希值所属的候选图像。

[0201] 在一种可能的实现方式中,为了能够预先构建出哈希表以及哈希表中的链表,该装置还可以包括:链表生成单元,用于通过如下方式生成所述目标哈希表中的链表:

[0202] 获取数据库中待处理的图像的哈希值;

[0203] 将所述图像的哈希值拆分所述预设数量个哈希值段,其中,不同的所述哈希值段在所述图像的哈希值中的位置顺序不同;

[0204] 对于每个哈希值段,定位与所述哈希值段的位置顺序对应的目标哈希表,并检测所述目标哈希表中是否存在用于记录所述哈希值段的链表;

[0205] 在所述目标哈希表中存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成的节点添加到所述用于记录所述哈希值段的链表的尾端;

[0206] 在所述目标哈希表中不存在用于记录所述哈希值段的链表的情况下,生成记录有所述哈希值段的节点,并将生成节点设定为所述目标哈希表中新增的链表的首个节点。

[0207] 在一种可能的实现方式中,以上装置的实施例中,该哈希检索单元可以包括:

[0208] 哈希发送子单元,用于将所述预设数量个连续的第一哈希值段发送给多台服务器,以便所述从服务器依据所述第一哈希值段在所述第一哈希值中的目标位置顺序,从所述从服务器负责检索的区域段内查询目标哈希表,所述区域段为数据库的存储空间中的部分存储区域;

[0209] 哈希检索子单元,用于通过所述从服务器从所述目标哈希表的多个哈希段中,检索出与所述第一哈希值段相同的第二哈希值段;

[0210] 所述图像匹配单元,包括:

[0211] 结果接收单元,用于接收所述从服务器返回的查询结果,所述查询结果至少包括:从所述第二哈希值段所属的候选图像中,检索出的与所述参考图像相似的目标图像的信息。

[0212] 进一步的,该装置还可以包括:

[0213] 故障监测单元,用于在监测到所述多个从服务器中存在运行异常的故障从服务器的情况下,确定所述故障从服务器当前负责检索的区域段;

[0214] 重分配单元,用于将所述故障从服务器对应的区域段分配给至少一个未出现运行异常的从服务器。

[0215] 参见图8,其示出了本申请实施例的图像检索方法所适用的服务器的一种组成结构示意图,该服务器可以为检索服务器。在图8中,该服务器800可以包括:处理器801和存储器802。当然,还可以包括通信接口803、输入单元804、显示器805和通信总线806。

[0216] 处理器801、存储器802、通信接口803、输入单元804、显示器805、均通过通信总线806完成相互间的通信。

[0217] 在本申请实施例中,该处理器801,可以为中央处理器(Central Processing Unit,CPU),特定应用集成电路(application-specific integrated circuit,ASIC),数字信号处理器(DSP)、专用集成电路(ASIC)、现成可编程门阵列(FPGA)或者其他可编程逻辑器件等。

[0218] 该处理器可以调用存储器802中存储的程序,具体的,可以处理器可以执行以上图2a、图3、图5以及图6中检索服务器或者主控服务器侧所执行的操作。

[0219] 存储器802中用于存放一个或者一个以上程序,程序可以包括程序代码,所述程序代码包括计算机操作指令,在本申请实施例中,该存储器中至少存储有用于实现以下功能的程序:

[0220] 获取待检索的参考图像的第一哈希值,所述第一哈希值用于表征所述参考图像的图像特征;

[0221] 将所述第一哈希值拆分成预设数量个连续的第一哈希值段,其中,不同的所述第一哈希值段在所述第一哈希值中的位置顺序不同;

[0222] 根据所述第一哈希值段的位置顺序,定位目标哈希表,并从所述目标哈希表的多个哈希值段中,检索出与所述第一哈希值段相同的第二哈希值段,所述目标哈希表用于存储图像的哈希值中,处于所述第一哈希值段的位置顺序上的哈希值段;

[0223] 从所述第二哈希值段所属的候选图像中,检索与所述参考图像相似的目标图像。

[0224] 在一种可能的实现方式中,该存储器802可包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、以及至少一个功能(比如声音播放功能、图像播放功能等)所需的应用程序等;存储数据区可存储根据计算机的使用过程中所创建的数据,比如,用户数据、用户访问数据以及音频数据等等。

[0225] 该通信接口803可以为通信模块的接口,如GSM模块的接口。

[0226] 当然,图8所示的计算机设备结构并不构成对本申请实施例中计算机设备的限定,在实际应用中计算机设备可以包括比图8所示的更多或更少的部件,或者组合某些部件。

[0227] 另一方面,本申请还提供了一种存储介质,该存储介质中存储有计算机程序,所述

计算机程序被处理器加载并执行时,实现如上任意一个实施例中所描述的图像检索方法。

[0228] .需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言,由于其与方法实施例基本相似,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0229] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0230] 对所公开的实施例的上述说明,使本领域技术人员能够实现或使用本发明。对这些实施例的多种修改对本领域技术人员来说将是显而易见的,本文中所定义的一般原理可以在不脱离本发明的精神或范围的情况下,在其它实施例中实现。因此,本发明将不会被限制于本文所示的这些实施例,而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

[0231] 以上仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

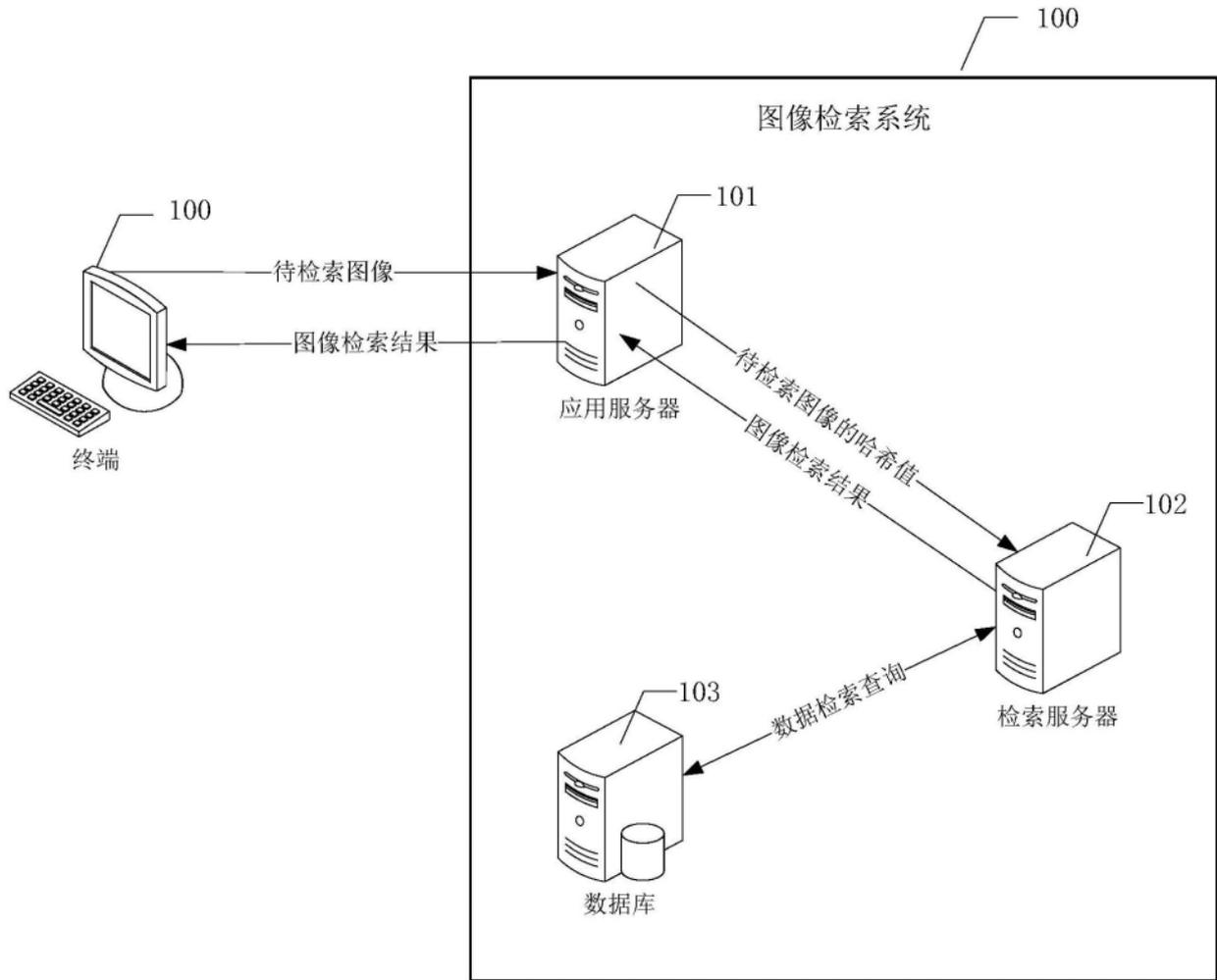


图1

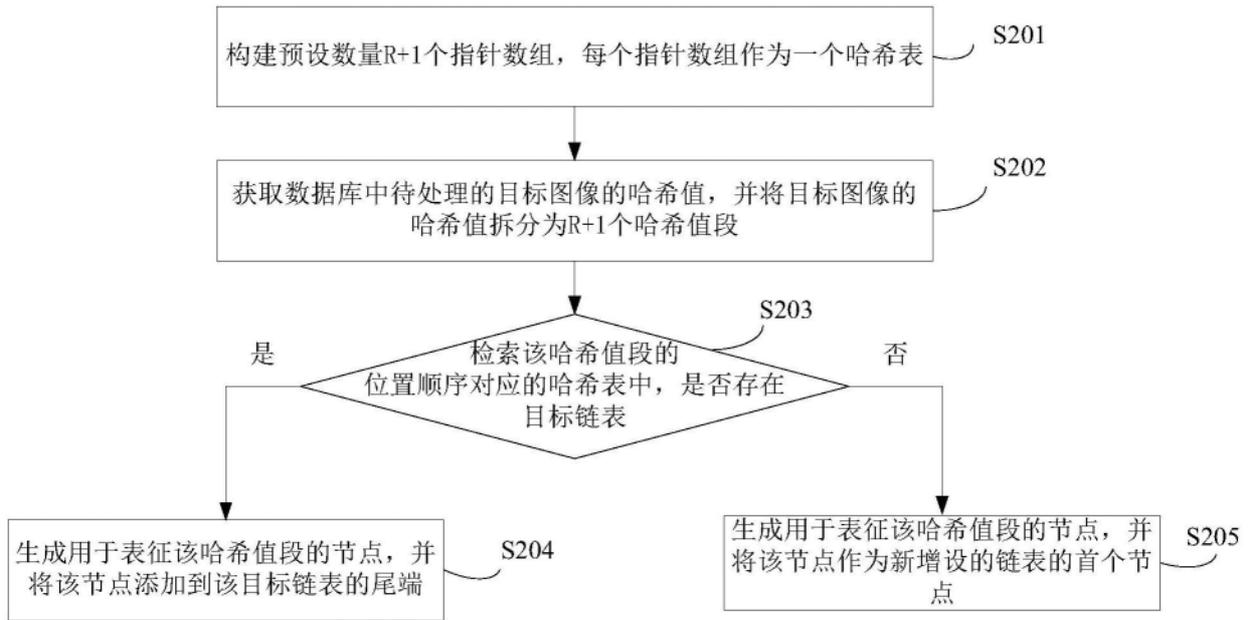


图2a

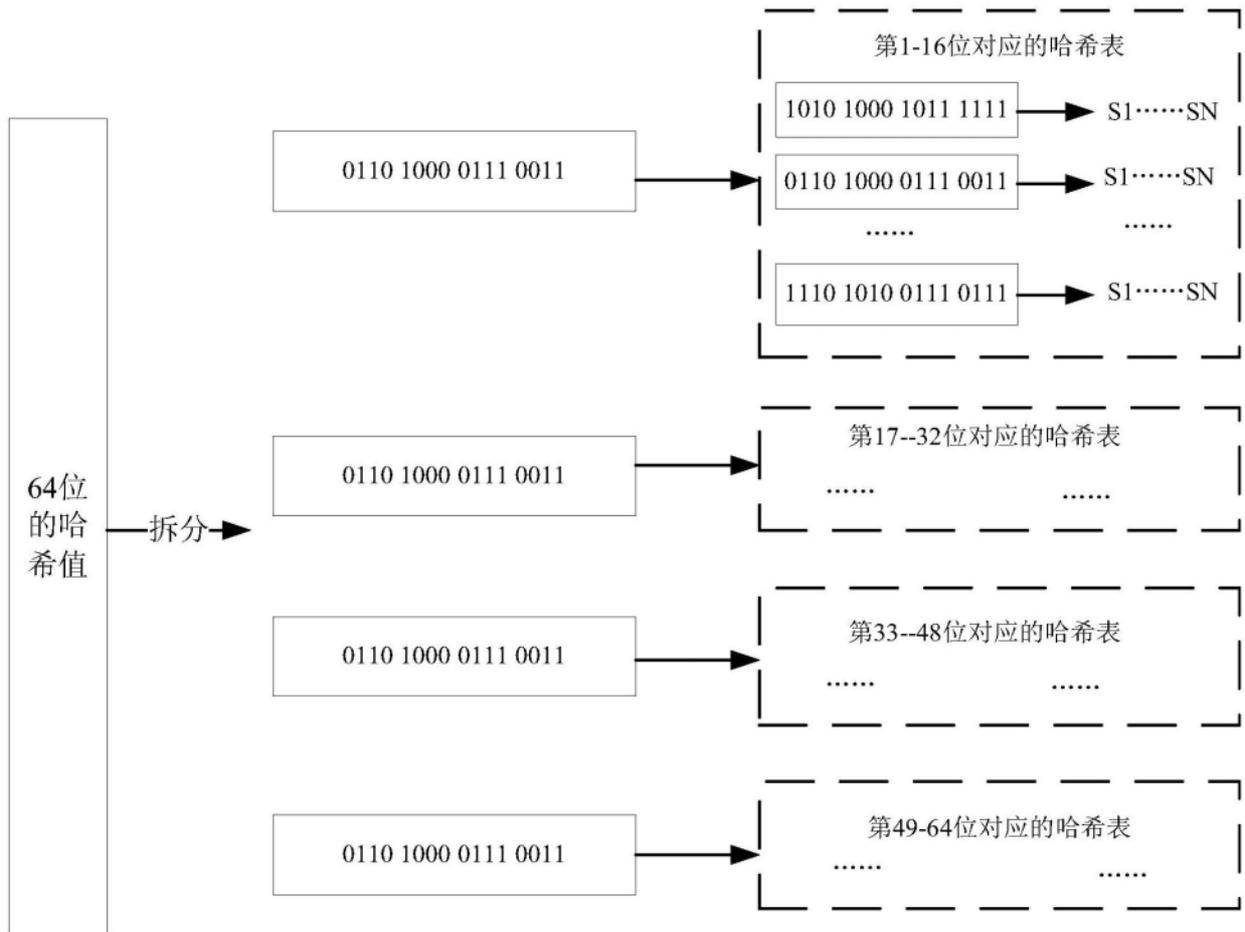


图2b

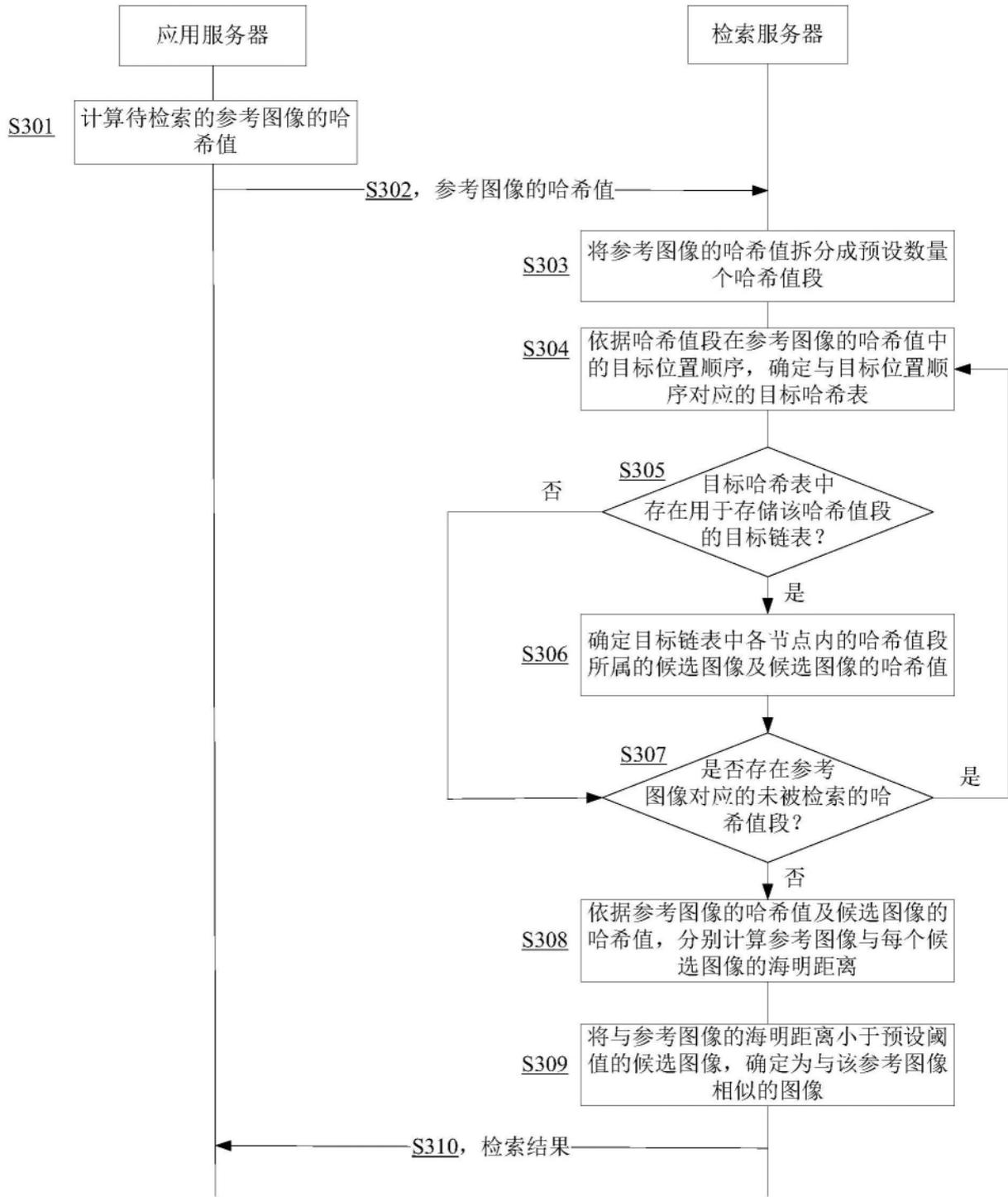


图3

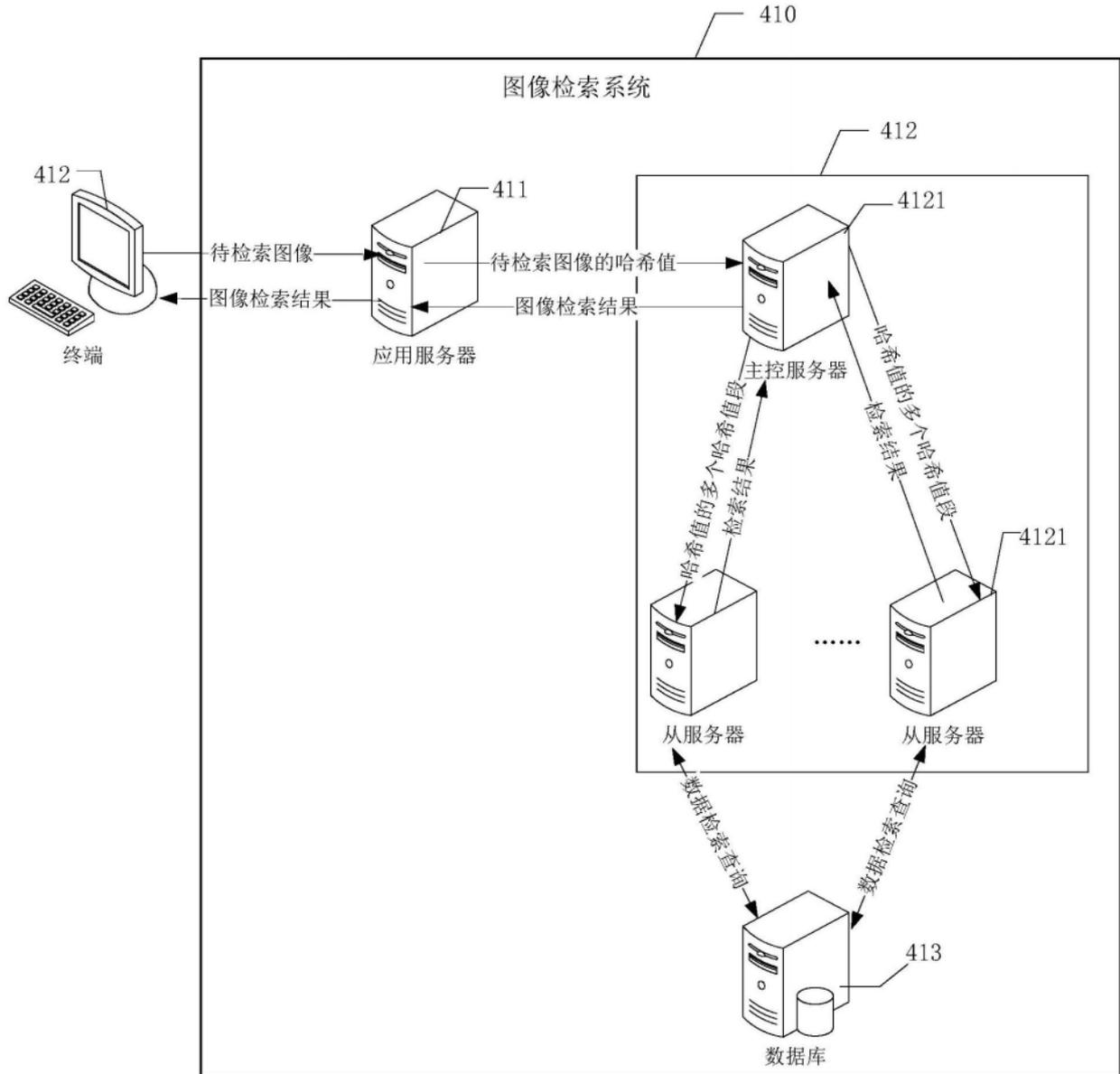


图4

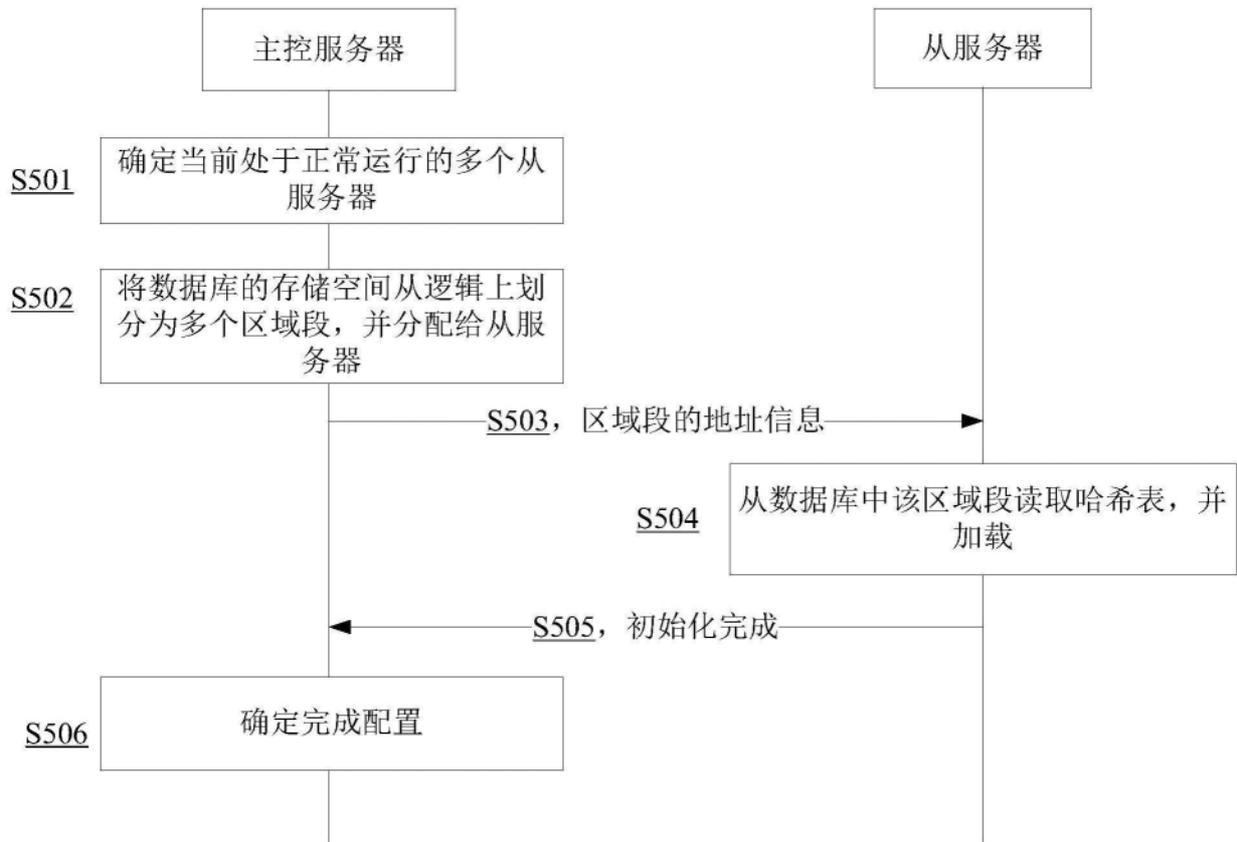


图5

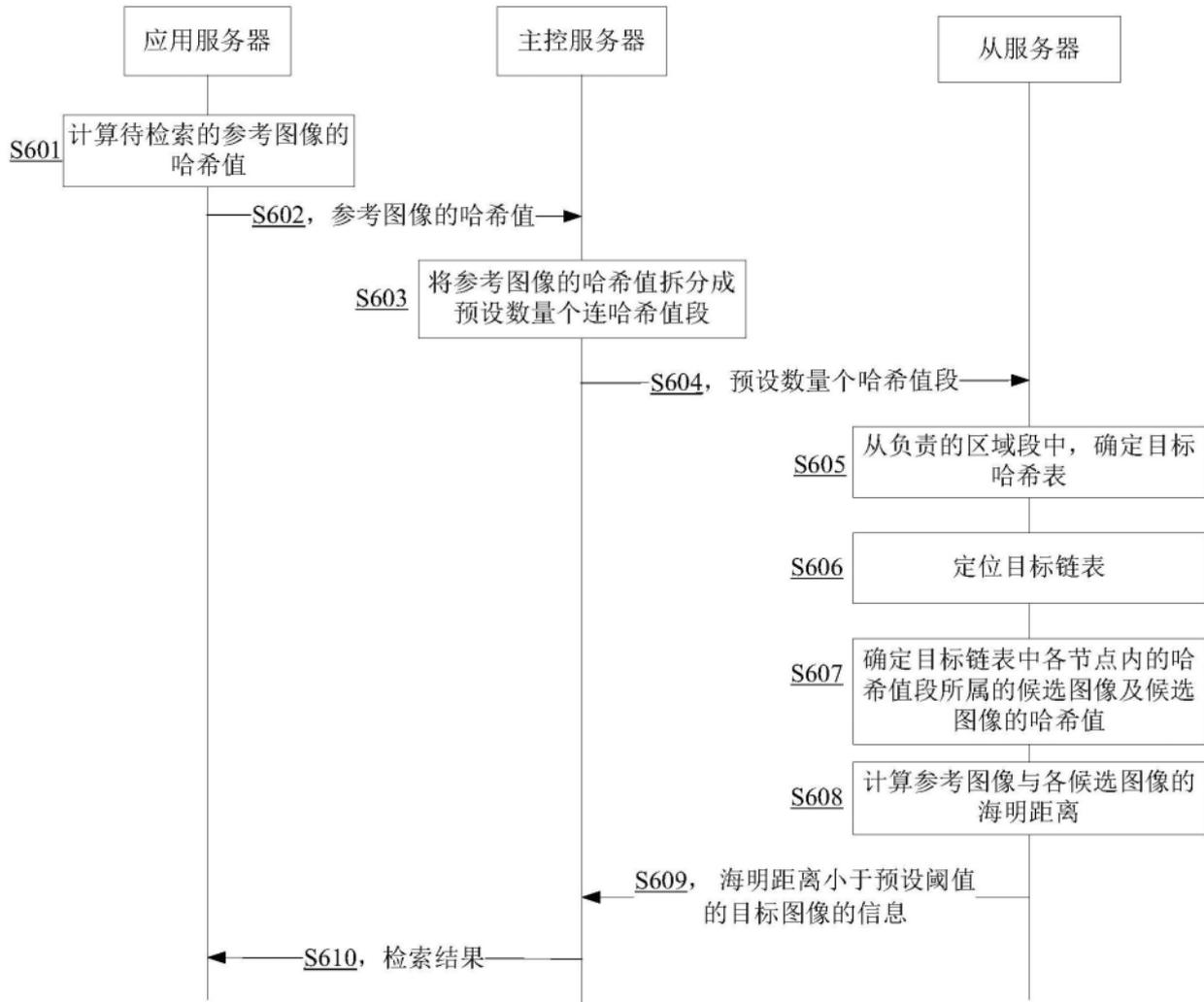


图6

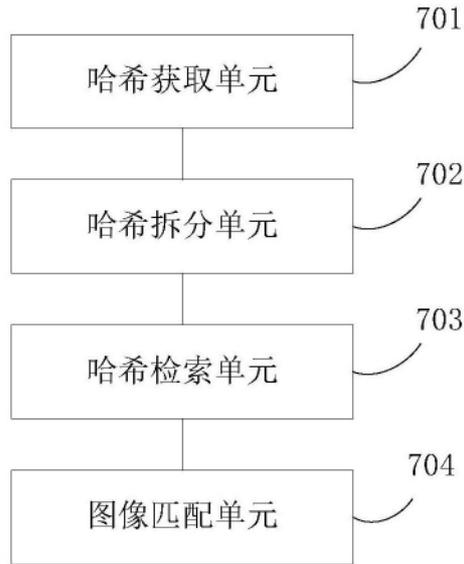


图7

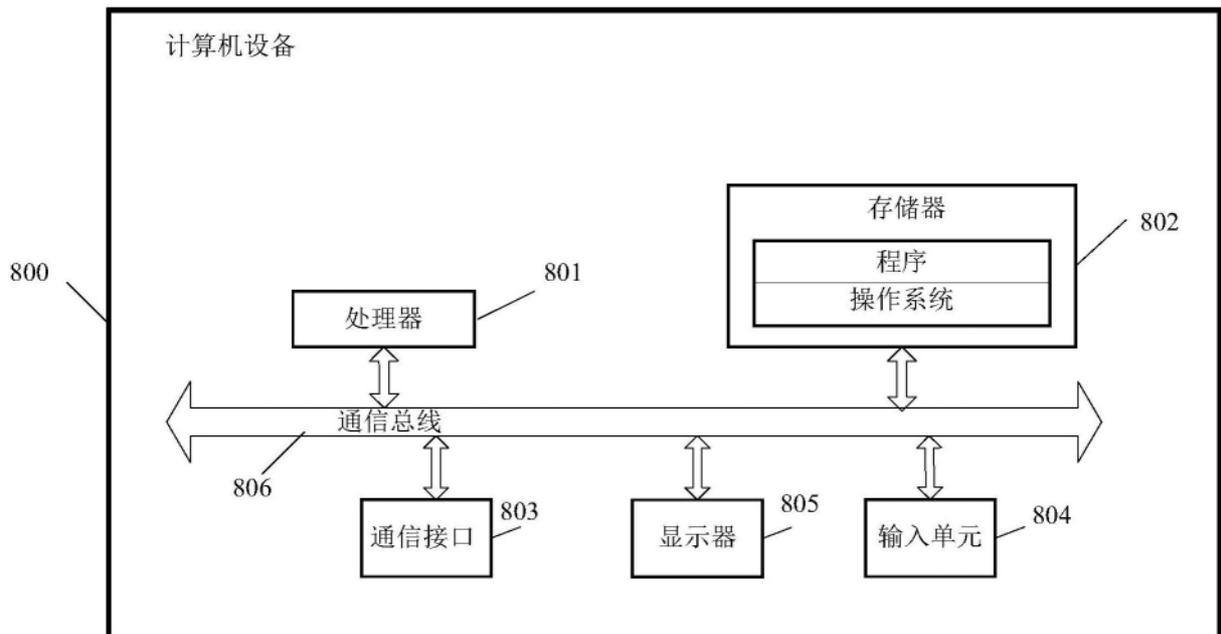


图8