



(12) 发明专利

(10) 授权公告号 CN 111368122 B

(45) 授权公告日 2022. 09. 30

(21) 申请号 202010092574.1

G06F 16/51 (2019.01)

(22) 申请日 2020.02.14

G06F 16/13 (2019.01)

G06F 16/174 (2019.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 111368122 A

(56) 对比文件

US 2015339316 A1, 2015.11.26

CN 109508400 A, 2019.03.22

US 2012159175 A1, 2012.06.21

CN 107729935 A, 2018.02.23

(43) 申请公布日 2020.07.03

(73) 专利权人 深圳壹账通智能科技有限公司

地址 518000 广东省深圳市前海深港合作区前湾一路1号A栋201室(入驻深圳市前海商务秘书有限公司)

审查员 夏雪

(72) 发明人 曾冰清

(74) 专利代理机构 北京中强智尚知识产权代理

有限公司 11448

专利代理师 黄耀威

(51) Int. Cl.

G06F 16/583 (2019.01)

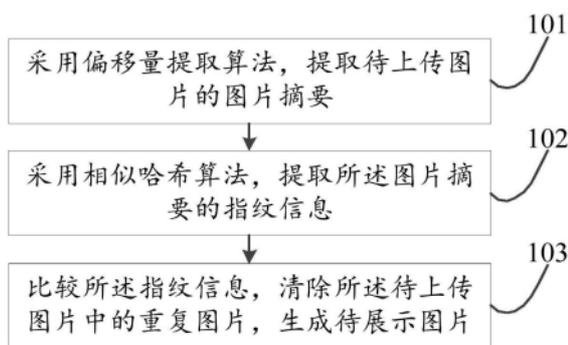
权利要求书2页 说明书8页 附图4页

(54) 发明名称

一种图片去重的方法及装置

(57) 摘要

本发明公开了一种图片去重的方法及装置,涉及图像处理技术领域,为解决现有技术中去重效率低的问题而发明。该方法主要包括:将待上传图片进行编号;采用偏移量提取算法,提取所述待上传图片的图片摘要;采用相似哈希算法,提取所述图片摘要的指纹信息;比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。本发明主要应用于在线交易平台图片上传的过程中。



1. 一种图片去重的方法,其特征在于,包括:

采用偏移量提取算法按照预置偏移量及预置行列调整函数,提取待上传图片的图片摘要;

采用相似哈希算法,提取所述图片摘要的指纹信息;

比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片;

所述采用偏移量提取算法按照预置偏移量及预置行列调整函数,提取待上传图片的图片摘要,包括:

按照预置偏移量,提取待上传图片的摘要信息,所述待上传图片和所述摘要信息都是矩阵向量;

按照预置行列调整函数,生成所述摘要信息的图片摘要,所述预置行列调整函数用于将矩阵向量转换为单行向量;

所述采用相似哈希算法,提取所述图片摘要的指纹信息,包括:

构建过滤表,所述过滤表是单行向量,所述过滤表中元素的初始值相同;

以所述图片摘要中的元素值为所述过滤表的有效编号位置,将所述有效编号位置的元素值替换为占位值,所述占位值与所述初始值不相同;

确定所述更新后的过滤表为所述指纹信息。

2. 如权利要求1所述的方法,其特征在于,所述按照预置行列调整函数,生成所述摘要信息的图片摘要之后,所述方法还包括:

按照预置更新规则,更新所述图片摘要的元素值,所述预置更新规则是将所述图片摘要中的元素值所属的编号位置与元素数值范围相乘获得乘积,将所述乘积与所述元素值进行加和运算,所述元素数值范围是指所述待上传图片的矩阵向量中单个数据位所能表示的最大数值。

3. 如权利要求1所述的方法,其特征在于,所述比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片,包括:

将所述待上传图片中的第一存放位图片的指纹信息,存放至指纹信息库;

依次比较所述指纹信息库中是否存在所述待上传图片中的其余存放位图片的指纹信息;

如果比较结果为不存在,则将所述待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库;

将所述指纹信息库中的指纹信息对应的所述待上传图片标记为待展示;

从所述待上传图片中选取具有所述待展示标记的待展示图片。

4. 如权利要求3所述的方法,其特征在于,所述将所述待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库之后,所述方法还包括:

保存所述指纹信息库,以便于判断二次上传图片是否为重复图片。

5. 如权利要求3所述的方法,其特征在于,所述比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片之后,所述方法还包括:

按照预置展示规则,展示所述待展示图片。

6. 一种图片去重的装置,其特征在于,包括:

第一提取模块,用于采用偏移量提取算法按照预置偏移量及预置行列调整函数,提取

待上传图片的图片摘要；

第二提取模块,用于采用相似哈希算法,提取所述图片摘要的指纹信息；

生成模块,用于比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片；

所述第一提取模块,包括：

提取单元,用于按照预置偏移量,提取待上传图片的摘要信息,所述待上传图片和所述摘要信息都是矩阵向量；

生成单元,用于按照预置行列调整函数,生成所述摘要信息的图片摘要,所述预置行列调整函数用于将矩阵向量转换为单行向量；

所述第二提取模块,包括：

构建单元,用于构建过滤表,所述过滤表是单行向量,所述过滤表中元素的初始值相同；

替换单元,用于以所述图片摘要中的元素值为所述过滤表的有效编号位置,将所述有效编号位置的元素值替换为占位值,所述占位值与所述初始值不相同；

确定单元,用于确定所述更新后的过滤表为所述指纹信息。

7. 一种计算机存储介质,其特征在于,所述计算机存储介质中存储有至少一可执行指令,所述可执行指令使处理器执行如权利要求1-5中任一项所述的图片去重的方法对应的操作。

8. 一种计算机设备,其特征在于,包括:处理器、存储器、通信接口和通信总线,所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信；

所述存储器用于存放至少一可执行指令,所述可执行指令使所述处理器执行如权利要求1-5中任一项所述的图片去重的方法对应的操作。

一种图片去重的方法及装置

技术领域

[0001] 本发明涉及图像处理技术领域,特别是涉及一种图片去重的方法及装置。

背景技术

[0002] 随着科技的发展和技术的进步,通过网络进行在线交易的人越来越多,在线交易平台的作用也越来越大。商户通常将产品图片上传至在线交易平台,在上传图片时可能会传入相同内容图片。相同内容图片的重复上传不仅给消费者带来不好的消费体验而且浪费系统的存储空间。

[0003] 现有技术中,采用通过感知哈希算法将待去重图片中的每幅图片生成图像指纹信息,再采用多组随机的哈希映射来构建图像哈希特征字典,从而去除重复图片。现有技术中将图片构成连通图和矩阵的方式,不仅运算复杂耗时太长,而且去重效率低。

发明内容

[0004] 有鉴于此,本发明提供一种图片去重的方法及装置,主要目的在于解决现有技术中去重效率低的问题。

[0005] 依据本发明一个方面,提供了一种图片去重的方法,包括:

[0006] 采用偏移量提取算法,提取待上传图片的图片摘要;

[0007] 采用相似哈希算法,提取所述图片摘要的指纹信息;

[0008] 比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0009] 依据本发明另一个方面,提供了一种图片去重的装置,包括:

[0010] 第一提取模块,用于采用偏移量提取算法,提取待上传图片的图片摘要;

[0011] 第二提取模块,用于采用相似哈希算法,提取所述图片摘要的指纹信息;

[0012] 生成模块,用于比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0013] 根据本发明的又一方面,提供了一种计算机存储介质,所述计算机存储介质中存储有至少一可执行指令,所述可执行指令使处理器执行如上述图片去重的方法对应的操作。

[0014] 根据本发明的再一方面,提供了一种计算机设备,包括:处理器、存储器、通信接口和通信总线,所述处理器、所述存储器和所述通信接口通过所述通信总线完成相互间的通信;

[0015] 所述存储器用于存放至少一可执行指令,所述可执行指令使所述处理器执行上述图片去重的方法对应的操作。

[0016] 借由上述技术方案,本发明实施例提供的技术方案至少具有下列优点:

[0017] 本发明提供了一种图片去重的方法及装置,首先将待上传图片进行编号,然后采用偏移量提取算法提取待上传图片的图片摘要,再采用相似哈希算法提取图片摘要的指纹信息,最后比较指纹信息清除待上传图片中的重复图片,生成待展示图片。与现有技术相

比,本发明实施例通过采用偏移量和相似哈希算法提取图片摘要的指纹信息,在比对重复图片的指纹信息时,由于比较操作是基于位运算的,所以单张重复图片的比对时间在秒级,提高比较去重效率。

[0018] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0019] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0020] 图1示出了本发明实施例提供的一种图片去重的方法流程图;

[0021] 图2示出了本发明实施例提供的另一种图片去重的方法流程图;

[0022] 图3示出了本发明实施例提供的一种图片去重的装置组成框图;

[0023] 图4示出了本发明实施例提供的另一种图片去重的装置组成框图;

[0024] 图5示出了本发明实施例提供的一种计算机设备的结构示意图。

具体实施方式

[0025] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0026] 商户通常将产品图片上传至在线交易平台,在上传图片时可能会传入相同内容图片。相同内容图片的重复上传不仅给消费者带来不好的消费体验而且浪费系统的存储空间。本发明实施例提供了一种图片去重的方法,如图1所示,该方法包括:

[0027] 101、采用偏移量提取算法,提取待上传图片的图片摘要。

[0028] 待上传图片是商户选定的准备上传至在线交易平台的产品图片。在线交易平台展示待上传图片之前,将待上传图片进行去重,以使得在线交易平台只展示不相同的产品图片。将待上传图片保存至磁盘,然后对待上传图片进行编号,再依次对每张待上传图片提取图片摘要。图片编号是待上传图片的图片标识,编号时可采用顺序编号、随机编号等方法,本发明实施例中对此不做限定。

[0029] 由于图片占用的内存较大,所以需要每张图片进行摘要提取,以减少去重过程中的数据计算量。偏移量提取算法,是指从表示图片的矩阵的首行首列为起始位置,查找与起始位置相距预置偏移量的摘要位置,提取起始位置和摘要位置对应的矩阵元素,并将矩阵元素作为图片摘要。待上传图片与图片摘要是一一对应的,图片摘要可采用与待上传图片相同的编号进行标识。

[0030] 102、采用相似哈希算法,提取所述图片摘要的指纹信息。

[0031] 依次对每次待上传图片提取图片摘要的指纹信息。相似哈希算法,是指把任意大小的图片摘要数据,通过散列算法变化成固定长度的指纹信息输出。散列算法,也就是将图

片摘要映射成指纹信息的映射方法。相同固定长度的指纹信息,便于后续比较指纹信息是否相同。图片摘要与指纹信息是一一对应的,指纹信息可以采用与图片摘要相同的编号进行标识,也就是待上传图片、待上传图片的图片摘要和待上传图片的指纹信息统一编号进行标识。

[0032] 103、比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0033] 按照编号的先后顺序,将第一编号位的指纹信息存放集合R中,将第二编号位的指纹信息与集合R中的指纹信息做比较,如果集合R中查找到相同的指纹信息则不做任何处理,如果集合R中未查找到相同的指纹信息则将第二编号位的指纹新存入集合R中。类似的,将其余编号位对应的指纹信息依次与集合R中的指纹信息做对比,如果在集合R中查找到相同的指纹信息则不做任何处理,如果在集合R中未查找到相同的指纹信息则将当前比对的指纹信息存入集合R中。通过前述方法,能够得到不包括重复指纹信息的集合R,集合R中的指纹信息对应的待上传图片即为待展示图片。

[0034] 本发明提供了一种图片去重的方法,首先将待上传图片进行编号,然后采用偏移量提取算法提取待上传图片的图片摘要,再采用相似哈希算法提取图片摘要的指纹信息,最后比较指纹信息清楚待上传图片中的重复图片,生成待展示图片。与现有技术相比,本发明实施例通过采用偏移量和相似哈希算法提取图片摘要的指纹信息,在比对重复图片的指纹信息时,由于比较操作是基于位运算的,所以单张重复图片的比对时间在秒级,提高比较去重效率。

[0035] 本发明实施例提供了另一种图片去重的方法,如图2所示,该方法包括:

[0036] 201、采用偏移量提取算法,提取待上传图片的图片摘要。

[0037] 由于图片占用的内存较大,所以需要每张图片进行摘要提取,以减少去重过程中的数据计算量。偏移量提取算法,是指从表示图片的矩阵的首行首列为起始位置,查找与起始位置相距预置偏移量的摘要位置,提取起始位置和摘要位置对应的矩阵元素,并将矩阵元素作为图片摘要。

[0038] 提取图片摘要具体包括:按照预置偏移量,提取待上传图片的摘要信息,所述待上传图片 and 所述摘要信息都是矩阵向量;按照预置行列调整函数,生成所述摘要信息的图片摘要,所述预置行列调整函数用于将矩阵向量转换为单行向量。其中,按照预置行列调整函数,生成所述摘要信息的图片摘要之后,所述方法还包括:按照预置更新规则,更新所述图片摘要的元素值,所述预置更新规则是将所述图片摘要中的元素值所属的编号位置与元素数值范围的乘积,与所述元素值进行加和运算,所述元素数值范围是指所述待上传图片的矩阵向量中单个数据位所能表示的最大数值。

[0039] 示例性的,一张图片可以用8行8列的图片矩阵表示,图片矩阵中的每个矩阵元素用1个字节表示。每个字节包括8比特,所能表示的最大数值数值为255。图片矩阵如下:

$$[0040] \begin{bmatrix} 0 & 254 & 253 & 196 & 195 & 194 & 121 & 120 \\ 0 & 0 & 8 & 255 & 198 & 1 & 0 & 0 \\ 18 & 110 & 112 & 192 & 220 & 0 & 0 & 0 \\ 0 & 0 & 8 & 255 & 196 & 1 & 0 & 0 \\ 1 & 2 & 9 & 255 & 254 & 100 & 0 & 0 \\ 0 & 3 & 10 & 255 & 223 & 99 & 0 & 0 \\ 0 & 0 & 8 & 255 & 190 & 1 & 255 & 0 \\ 0 & 0 & 110 & 112 & 196 & 110 & 0 & 0 \end{bmatrix}$$

[0041] 假设预置偏移量为两个字节,也就是每两个自己取前一个字节,提取摘要信息。也就是提取图片矩阵中的偶数行和偶数列,得到4行4列的摘要信息的矩阵,矩阵如下:

$$[0042] \begin{bmatrix} 0 & 253 & 195 & 121 \\ 18 & 112 & 220 & 0 \\ 1 & 9 & 254 & 0 \\ 0 & 8 & 190 & 155 \end{bmatrix}$$

[0043] 将摘要信息转换为单行的图片摘要,与摘要信息对应的图片摘要为:(0,253,195,121,18,112,220,0,1,9,254,0,0,8,190,255)。为了增加不同位置的元素值区分度,按照预置更新规则更新图片摘要中的元素值。首先计算元素数值范围,由于待上传图片的矩阵向量中单个数据位为8位,其能表示的最大数据是2的8次方也就是255。然后将所述图片摘要中的元素值所属的编号位置与元素数值范围的乘积,与所述元素值进行加和运算,图片摘要为单行向量,行向量的位置编号通常从0开始计数,上述示例中图片摘要的元素值所述的编号位置按照顺序为0至15。将元素值的编号位置与255相乘,再与该元素值进行加和,获取更新后的元素值。上述示例中,更新元素值分别为0,253+255,195+255*2,121+255*3,18+255*4,112+255*5,220+255*6,0+255*7,1+255*8,9+255*9,254+255*10,0+255*11,0+255*12,8+255*13,190+255*14,255+255*15。图片摘要为(0,508,702,886,1038,1387,1750,1785,2041,2304,2804,2805,3060,3323,3760,4080)。

[0044] 202、采用相似哈希算法,提取所述图片摘要的指纹信息。

[0045] 提取指纹信息具体包括:构建过滤表,所述过滤表是单行向量,所述过滤表中元素的初始值相同;以所述图片摘要中的元素值为所述过滤表的有效编号位置,将所述有效编号位置的元素值替换为占位值,所述占位值与所述初始值不相同;确定所述更新后的过滤表为所述指纹信息。

[0046] 以步骤201中的示例为例,继续说明本步骤。在图片摘要中最后一位元素值为255+255*15=4080,由于图片摘要的元素值计算规则以及元素值范围可知,更新后图片摘要中可能存在的最大元素值为4080。一方面为了能够表示所有的数值,另一方面为了能够用二进制标识,所以选取大于最大元素值且是2的整数次方的数值作为过滤表的数据位数,满足条件的数据为4096。过滤表包括4096个元素,范围是0-4095,将其初始值都设置为0。根据图片摘要中的元素值,确定过滤表中的有效编号位置分别为0,508,702,886,1038,1387,1750,1785,2041,2304,2804,2805,3060,3323,3760,4080。将过滤表中的有效编号位置的元素值替换为占位值,占位值与初始值不同,可以选取1为占位值。由于计算过程中机器识

别的数据为二进制数,所以初始值和占位值选用0和1能够提高数据的计算速度。

[0047] 203、比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0048] 生成待展示图片,具体包括:将所述待上传图片中的第一存放位图片的指纹信息,存放至指纹信息库;依次比较所述指纹信息库中是否存在所述待上传图片中的其余存放位图片的指纹信息;如果所述比较结果为不存在,则将所述待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库;将所述指纹信息库中的指纹信息对应的所述待上传图片标记为待展示;从所述待上传图片中选取具有所述待展示标记的待展示图片。

[0049] 在比较指纹信息时,依次比较过滤表中的各个数据位中的元素值是否相同即可,在比较过程中可采用同或运算,或者采用异或运算。比较过程基于位运算,判断时间在秒级,耗时短,十分高效。对于8*8的图片,该过滤器的内存容量大小只有4096比特,只需要512个字节即0.5KB内存。几乎不怎么消耗计算机内存。这种图片摘要提取和指纹比对的方法最终将用户的重复上传的图片进行去重,从而降低内存消耗,给用户更好的体验。

[0050] 经前述运算,在指纹信息库中存放的指纹信息,都是不完全相同的,指纹信息对应的待上传图片都不是重复图片。将不重复的图片标记为待展示,然后将选取带展示标记的带展示图片。

[0051] 将待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库之后,还包括:保存所述指纹信息库,以便于判断二次上传图片是否为重复图片。商户往往需要多次上传图片,保存第一次去重判断过程中中的指纹信息库,在后续二次上传图片时,根据原上传图片的指纹信息库,判断二次上传的图片是否与之前上传的图片是重复的,以使得每次上传的图片与当前已上传的所有图片进行比较。当然,商户删除在线平台上已展示的图片后,也相应的删除指纹信息库中的对应的指纹信息。

[0052] 204、按照预置展示规则,展示所述待展示图片。

[0053] 预置展示规则,可以为根据待展示图片的保存顺序展示,可以为响应商户的操作分类显示,在本发明实施例中对此不做限定。根据预置展示规则,在在线平台的展示区展示待展示图片。

[0054] 本发明提供了一种图片去重的方法,首先将待上传图片进行编号,然后采用偏移量提取算法提取待上传图片的图片摘要,再采用相似哈希算法提取图片摘要的指纹信息,最后比较指纹信息清楚待上传图片中的重复图片,生成待展示图片。与现有技术相比,本发明实施例通过采用偏移量和相似哈希算法提取图片摘要的指纹信息,在比对重复图片的指纹信息时,由于比较操作是基于位运算的,所以单张重复图片的比对时间在秒级,提高比较去重效率。

[0055] 进一步的,作为对上述图1所示方法的实现,本发明实施例提供了一种图片去重的装置,如图3所示,该装置包括:

[0056] 第一提取模块31,用于采用偏移量提取算法,提取待上传图片的图片摘要;

[0057] 第二提取模块32,用于采用相似哈希算法,提取所述图片摘要的指纹信息;

[0058] 生成模块33,用于比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0059] 本发明提供了一种图片去重的装置,首先将待上传图片进行编号,然后采用偏移量提取算法提取待上传图片的图片摘要,再采用相似哈希算法提取图片摘要的指纹信息,

最后比较指纹信息清楚待上传图片中的重复图片,生成待展示图片。与现有技术相比,本发明实施例通过采用偏移量和相似哈希算法提取图片摘要的指纹信息,在比对重复图片的指纹信息时,由于比较操作是基于位运算的,所以单张重复图片的比对时间在秒级,提高比较去重效率。

[0060] 进一步的,作为对上述图2所示方法的实现,本发明实施例提供了另一种图片去重的装置,如图4所示,该装置包括:

[0061] 第一提取模块41,用于采用偏移量提取算法,提取待上传图片的图片摘要;

[0062] 第二提取模块42,用于采用相似哈希算法,提取所述图片摘要的指纹信息;

[0063] 生成模块43,用于比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0064] 进一步的,所述第一提取模块41,包括:

[0065] 提取单元411,用于按照预置偏移量,提取待上传图片的摘要信息,所述待上传图片 and 所述摘要信息都是矩阵向量;

[0066] 生成单元412,用于按照预置行列调整函数,生成所述摘要信息的图片摘要,所述预置行列调整函数用于将矩阵向量转换为单行向量。

[0067] 进一步的,所述装置还包括:

[0068] 更新单元413,用于所述按照预置行列调整函数,生成所述摘要信息的图片摘要之后,按照预置更新规则,更新所述图片摘要的元素值,所述预置更新规则是将所述图片摘要中的元素值所属的编号位置与元素数值范围的乘积,与所述元素值进行加和运算,所述元素数值范围是指所述待上传图片的矩阵向量中单个数据位所能表示的最大数值。

[0069] 进一步的,所述第二提取模块42,包括:

[0070] 构建单元421,用于构建过滤表,所述过滤表是单行向量,所述过滤表中元素的初始值相同;

[0071] 替换单元422,用于以所述图片摘要中的元素值为所述过滤表的有效编号位置,将所述有效编号位置的元素值替换为占位值,所述占位值与所述初始值不相同;

[0072] 确定单元423,用于确定所述更新后的过滤表为所述指纹信息。

[0073] 进一步的,所述生成模块43,包括:

[0074] 存放单元431,用于将所述待上传图片中的第一存放位图片的指纹信息,存放至指纹信息库;

[0075] 比较单元432,用于依次比较所述指纹信息库中是否存在所述待上传图片中的其余存放位图片的指纹信息;

[0076] 所述存放单元431,还用于如果所述比较结果为不存在,则将所述待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库;

[0077] 标记单元433,用于将所述指纹信息库中的指纹信息对应的所述待上传图片标记为待展示;

[0078] 选取单元434,用于从所述待上传图片中选取具有所述待展示标记的待展示图片。

[0079] 进一步的,所述装置还包括:

[0080] 保存单元435,用于所述将所述待上传图片中的其余存放位图片的指纹信息存放至所述指纹信息库之后,保存所述指纹信息库,以便于判断二次上传图片是否为重复图片。

[0081] 进一步的,所述装置还包括:

[0082] 展示模块44,用于所述比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片之后,按照预置展示规则,展示所述待展示图片。

[0083] 本发明提供了一种图片去重的装置,首先将待上传图片进行编号,然后采用偏移量提取算法提取待上传图片的图片摘要,再采用相似哈希算法提取图片摘要的指纹信息,最后比较指纹信息清除待上传图片中的重复图片,生成待展示图片。与现有技术相比,本发明实施例通过采用偏移量和相似哈希算法提取图片摘要的指纹信息,在比对重复图片的指纹信息时,由于比较操作是基于位运算的,所以单张重复图片的比对时间在秒级,提高比较去重效率。

[0084] 根据本发明一个实施例提供了一种计算机存储介质,所述计算机存储介质存储有至少一可执行指令,该计算机可执行指令可执行上述任意方法实施例中的图片去重的方法。

[0085] 图5示出了根据本发明一个实施例提供的一种计算机设备的结构示意图,本发明具体实施例并不对计算机设备的具体实现做限定。

[0086] 如图5所示,该计算机设备可以包括:处理器(processor)502、通信接口(Communications Interface)504、存储器(memory)506、以及通信总线508。

[0087] 其中:处理器502、通信接口504、以及存储器506通过通信总线508完成相互间的通信。

[0088] 通信接口504,用于与其它设备比如客户端或其它服务器等的网元通信。

[0089] 处理器502,用于执行程序510,具体可以执行上述图片去重的方法实施例中的相关步骤。

[0090] 具体地,程序510可以包括程序代码,该程序代码包括计算机操作指令。

[0091] 处理器502可能是中央处理器CPU,或者是特定集成电路ASIC(Application Specific Integrated Circuit),或者是被配置成实施本发明实施例的一个或多个集成电路。计算机设备包括的一个或多个处理器,可以是同一类型的处理器,如一个或多个CPU;也可以是不同类型的处理器,如一个或多个CPU以及一个或多个ASIC。

[0092] 存储器506,用于存放程序510。存储器506可能包含高速RAM存储器,也可能还包括非易失性存储器(non-volatile memory),例如至少一个磁盘存储器。

[0093] 程序510具体可以用于使得处理器502执行以下操作:

[0094] 采用偏移量提取算法,提取待上传图片的图片摘要;

[0095] 采用相似哈希算法,提取所述图片摘要的指纹信息;

[0096] 比较所述指纹信息,清除所述待上传图片中的重复图片,生成待展示图片。

[0097] 显然,本领域的技术人员应该明白,上述的本发明的各模块或各步骤可以用通用的计算装置来实现,它们可以集中在单个的计算装置上,或者分布在多个计算装置所组成的网络上,可选地,它们可以用计算装置可执行的程序代码来实现,从而,可以将它们存储在存储装置中由计算装置来执行,并且在某些情况下,可以以不同于此处的顺序执行所示出或描述的步骤,或者将它们分别制作成各个集成电路模块,或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样,本发明不限制于任何特定的硬件和软件结合。

[0098] 以上所述仅为本发明的优选实施例而已,并不用于限制本发明,对于本领域的技

术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包括在本发明的保护范围之内。

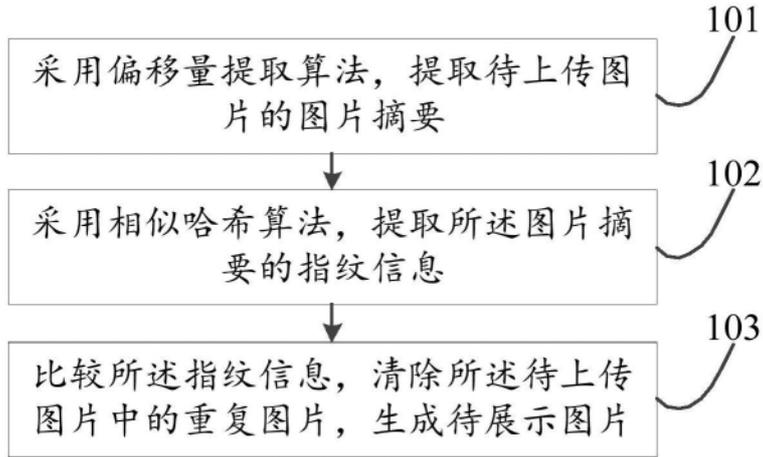


图1

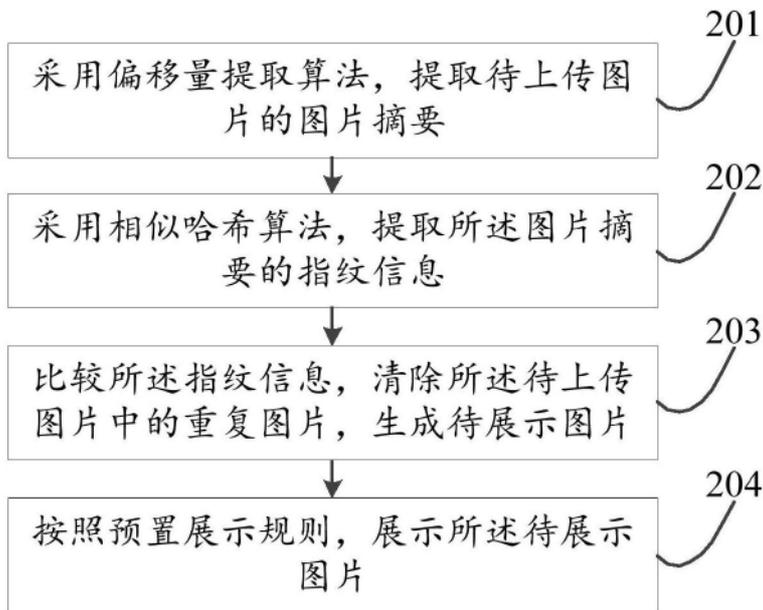


图2

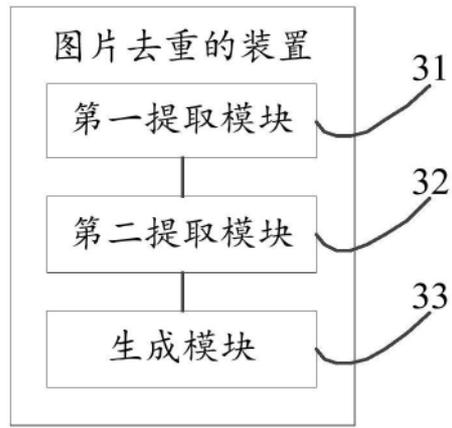


图3

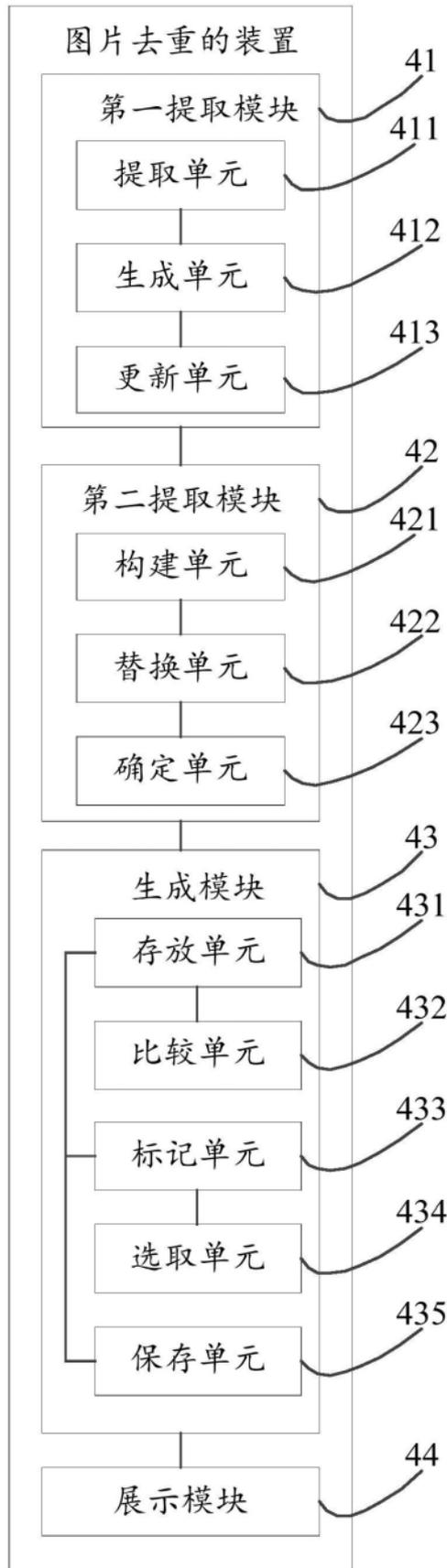


图4

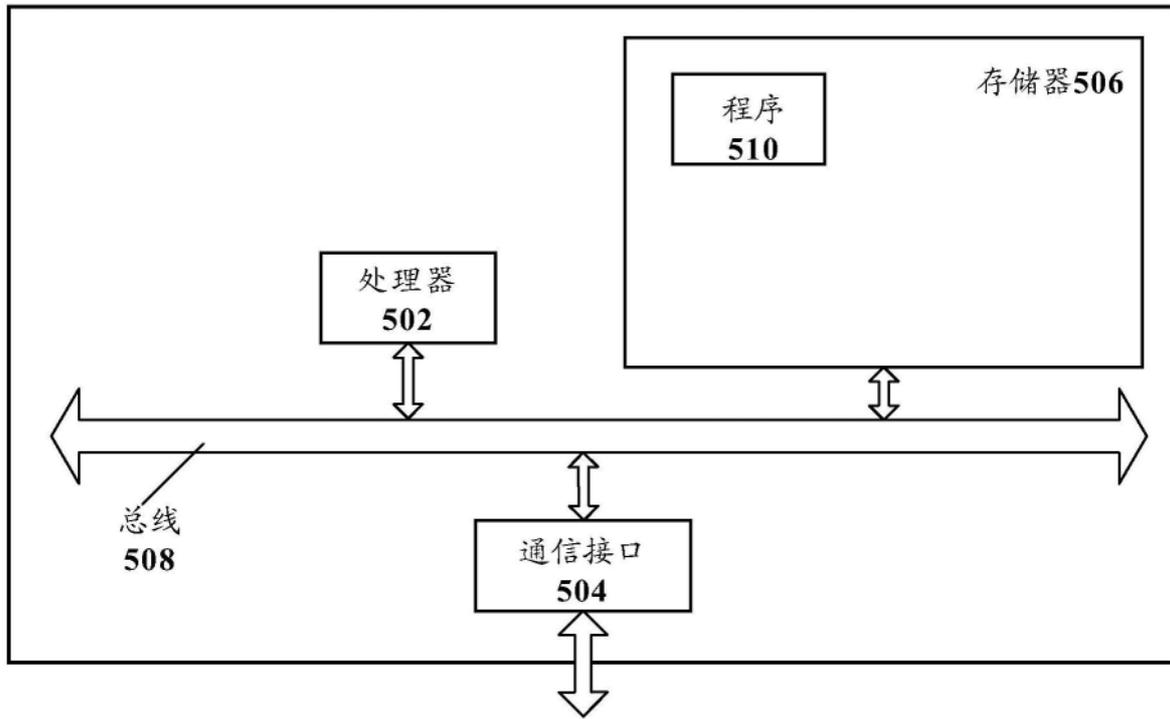


图5