



(12) 发明专利申请

(10) 申请公布号 CN 114637870 A

(43) 申请公布日 2022.06.17

(21) 申请号 202210246863.1

(22) 申请日 2022.03.14

(71) 申请人 重庆大学

地址 404100 重庆市沙坪坝区沙正街174号

(72) 发明人 谭玉娟 肖丹 晏志超 江泓

(74) 专利代理机构 广州三环专利商标代理有限公司

公司 44202

专利代理师 廖慧琪

(51) Int. Cl.

G06F 16/51 (2019.01)

G06T 1/00 (2006.01)

G06T 7/11 (2017.01)

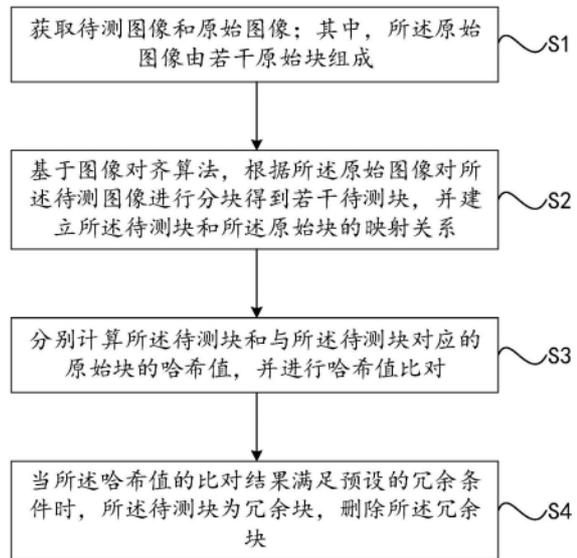
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种图像数据处理方法、装置、设备及存储介质

(57) 摘要

本发明公开一种图像数据处理方法、装置、设备及存储介质,包括:获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。本发明实施例通过对待测图像进行分块并建立待测块和原始图像的原始块之间的映射关系,根据待测块和对应的原始块的哈希值比对结果来确定并删除冗余块,实现了感知冗余内容的删除,优化了图像数据的存储。



1. 一种图像数据处理方法,其特征在于,包括:
获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;
基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;
分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;
当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。
2. 如权利要求1所述的图像数据处理方法,其特征在于,还包括:
当所述哈希值的比对结果不满足所述冗余条件时,所述待测块为非冗余块,将所述非冗余块保存至图像存储器;其中,所述冗余条件为:
所述待测块的哈希值和与所述待测块对应的原始块的哈希值之间的汉明距离小于预设距离阈值。
3. 如权利要求1所述的图像数据处理方法,其特征在于,所述原始图像通过以下方式获取:
提取所述待测图像的标识信息;
从图像存储器中选取标识信息与所述待测图像的标识信息相同的图像作为原始图像。
4. 如权利要求3所述的图像数据处理方法,其特征在于,在所述原始图像存入所述图像存储器之前,还包括:
基于预设的图像分块规则,将所述原始图像分块得到若干原始块;
获取原始标识信息;
将所述原始标识信息添加到每一所述原始块中。
5. 如权利要求4所述的图像数据处理方法,其特征在于,所述原始标识信息通过以下方式生成:获取所述原始图像的描述信息,根据所述描述信息生成原始标识信息;其中,所述描述信息包括与所述原始图像生成相关的用户、设备、应用程序、日期、位置或分辨率中的至少一种。
6. 如权利要求2所述的图像数据处理方法,其特征在于,在所述删除所述冗余块之前,还包括:
提取所述冗余块的图像信息,提取与所述冗余块对应的原始块的图像信息;
根据所述冗余块的图像信息、与所述冗余块对应的原始块的图像信息,得到所述冗余块的图像操作信息,并将所述图像操作信息存储于图像描述存储器中;
所述将所述非冗余块保存至图像存储器,具体包括:
将所述非冗余块附近的图像内容添加到所述非冗余块中,将添加所述图像内容后的所述非冗余块保存至所述图像存储器。
7. 如权利要求6所述的图像数据处理方法,其特征在于,在所述删除所述冗余块之后,还包括:
响应于图像恢复指令,根据所述图像操作信息和所述原始图像生成冗余图像;
根据所述冗余图像和所述非冗余块,恢复所述待测图像。
8. 一种图像数据处理装置,其特征在于,包括:
图像获取模块,用于获取待测图像和原始图像;其中,所述原始图像由若干原始块组

成；

图像分块模块,用于基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系；

图像比对模块,用于分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对；

图像删除模块,用于当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。

9. 一种图像数据处理设备,其特征在于,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至7中任意一项所述的图像数据处理方法。

10. 一种计算机可读存储介质,其特征在于,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如权利要求1至7中任意一项所述的图像数据处理方法。

一种图像数据处理方法、装置、设备及存储介质

技术领域

[0001] 本发明涉及互联网技术领域,尤其涉及一种图像数据处理方法、装置、设备及存储介质。

背景技术

[0002] 随着计算机技术的发展,无处不在的智能移动设备可以随时随地捕获电子图像并在社交媒体应用程序上广泛传播图像。在图像的传播和共享过程中,很容易从单个原始图像中衍生出大量相似的图像,首先,用户总是倾向于修改图像以达到自己的期望或表达自己的感受,其次,一张图像可以根据不同压缩级别被几十种流行的压缩算法压缩,从而产生大量感知内容相同但码流内容完全不同的压缩副本,此外,主流的以图像处理或以人工智能算法驱动的图片编辑工具也可能产生大量高度相似的图像,这些图像共享大量冗余的感知内容,但比特流完全不同。而现有的重复数据删除工具是基于比特流的,不能有效地识别具有不同比特流的感知冗余图像。

发明内容

[0003] 本发明实施例的目的是提供一种图像数据处理方法、装置、设备及存储介质,能够通过对待测图像进行分块并建立待测块和原始图像的原始块之间的映射关系,根据待测块和对应的原始块的哈希值比对结果来确定并删除冗余块,实现了感知冗余内容的删除,优化了图像数据的存储。

[0004] 为实现上述目的,本发明实施例提供了一种图像数据处理方法,包括:

[0005] 获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;

[0006] 基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;

[0007] 分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;

[0008] 当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。

[0009] 作为上述方案的改进,还包括:

[0010] 当所述哈希值的比对结果不满足所述冗余条件时,所述待测块为非冗余块,将所述非冗余块保存至图像存储器;其中,所述冗余条件为:

[0011] 所述待测块的哈希值和与所述待测块对应的原始块的哈希值之间的汉明距离小于预设距离阈值。

[0012] 作为上述方案的改进,所述原始图像通过以下方式获取:

[0013] 提取所述待测图像的标识信息;

[0014] 从图像存储器中选取标识信息与所述待测图像的标识信息相同的图像作为原始图像。

- [0015] 作为上述方案的改进,在所述原始图像存入所述图像存储器之前,还包括:
- [0016] 基于预设的图像分块规则,将所述原始图像分块得到若干原始块;
- [0017] 获取原始标识信息;
- [0018] 将所述原始标识信息添加到每一所述原始块中。
- [0019] 作为上述方案的改进,所述原始标识信息通过以下方式生成:获取所述原始图像的描述信息,根据所述描述信息生成原始标识信息;其中,所述描述信息包括与所述原始图像生成相关的用户、设备、应用程序、日期、位置或分辨率中的至少一种。
- [0020] 作为上述方案的改进,在所述删除所述冗余块之前,还包括:
- [0021] 提取所述冗余块的图像信息,提取与所述冗余块对应的原始块的图像信息;
- [0022] 根据所述冗余块的图像信息、与所述冗余块对应的原始块的图像信息,得到所述冗余块的图像操作信息,并将所述图像操作信息存储于图像描述存储器中;
- [0023] 所述将所述非冗余块保存至图像存储器,具体包括:
- [0024] 将所述非冗余块附近的图像内容添加到所述非冗余块中,将添加所述图像内容后的所述非冗余块保存至所述图像存储器。
- [0025] 作为上述方案的改进,在所述删除所述冗余块之后,还包括:
- [0026] 响应于图像恢复指令,根据所述图像操作信息和所述原始图像生成冗余图像;
- [0027] 根据所述冗余图像和所述非冗余块,恢复所述待测图像。
- [0028] 为实现上述目的,本发明实施例还提供了一种图像数据处理装置,包括:
- [0029] 图像获取模块,用于获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;
- [0030] 图像分块模块,用于基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;
- [0031] 图像比对模块,用于分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;
- [0032] 图像删除模块,用于当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。
- [0033] 为实现上述目的,本发明实施例还提供了一种图像数据处理设备,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如上述任一实施例所述的图像数据处理方法。
- [0034] 为实现上述目的,本发明实施例还提供了一种计算机可读存储介质,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如上述任一实施例所述的图像数据处理方法。
- [0035] 与现有技术相比,本发明实施例公开的图像数据处理方法、装置、设备和计算机可读存储介质,首先,获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;接着,基于图像对齐算法,通过根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;然后分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;最后,当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。本发明实施例能够通过对待测图像进行分块并建立待测块和原始图像的原始块之间的映射关系,根据待测块和对应的原始块的哈希

值比对结果来确定并删除冗余块,实现了感知冗余内容的删除,优化了图像数据的存储。

附图说明

- [0036] 图1是本发明一实施例提供的一种图像数据处理方法的流程图;
- [0037] 图2是本发明一实施例提供的一种原始图像和待测图像的分块示意图;
- [0038] 图3是本发明一实施例提供的一种图像存储空间要求柱形图;
- [0039] 图4是本发明一实施例提供的一种图像数据处理的结构框图。

具体实施方式

[0040] 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。

[0041] 参见图1,是本发明实施例提供的一种图像数据处理方法的流程图,所述图像数据处理方法包括步骤S1~S4:

[0042] S1、获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;

[0043] S2、基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;

[0044] S3、分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;

[0045] S4、当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。

[0046] 值得说明的是,所述图像数据处理方法可以通过用户端执行,所述用户端可以是电脑、手机、平板等用户终端设备。

[0047] 在步骤S1中,作为举例,可以通过响应于图像去重指令获取待测图像和原始图像;图像去重指令可以是用户输入的,输入方式可以是键盘输入、鼠标输入、触屏输入,在此不作限定;图像去重指令也可以是预先存储在用户终端,可以是当检测到新的图像生成或者接收到新的图像时,触发去重指令,也可以是预先设定去重程序的启动周期,每隔预设的时间进行一次去重操作。

[0048] 在步骤S1中,值得说明的是,待测图像是原始图像经过图像修改操作得到的,原始图像是由若干原始块组成;因此,在步骤S2中,作为举例,基于图像对齐算法,如RANSAC (random sample consensus) 算法,分别对待测图像和原始图像进行特征提取并进行特征匹配,以求解出对齐矩阵,结合原始图像由若干原始块组成的特性,从而实现待测图像的分块,以得到若干待测块,进一步建立起待测块和原始块的映射关系。

[0049] 在步骤S3中,由于对图像进行了分块处理,因此,不必以整个图像作为重复数据对象的分析,而是以更小的分析单元——待测块作为重复数据对象进行分析,采用感知哈希算法来计算待测块和原始块的哈希值,并基于待测块和原始块的映射关系,将待测块和与待测块对应的原始块进行哈希值的比对。示例性的,采用余弦感知哈希 (pHash) 计算待测块和原始块的哈希值,将每一待测块的哈希值和与待测块对应的原始块的哈希值进行汉明距

离的计算。

[0050] 在步骤S4中,具体地,分别对每一待测块与对应的原始块的哈希值的比对结果进行分析,判断每一比对结果是否满足预设的冗余条件,当比对结果满足预设的冗余条件时,说明该待测块与其对应的原始块为相似块,即该待测块为冗余块,因此将冗余块进行删除,以达到优化存储空间的目的。

[0051] 与现有技术相比,本发明实施例能够基于图像对齐算法,将获取的待测图像和原始图像进行对齐;结合原始图像的分块情况,对待测图像进行分块处理得到待测块,并建立其原始图像的原始块和待测图像的待测块之间的一一对应的映射关系;基于映射关系进行待测块和原始块之间的哈希值比对,进而根据哈希值的比对结果来确定冗余块,以用于删除冗余块,实现了块级的感知冗余内容的删除,降低了直接以待测图像作为冗余检测对象来检测冗余的计算量和计算延迟,同时也能针对整体相似性较小的两个图像中的冗余部分进行有效删除,减少存储空间的浪费,优化了图像的数据存储。

[0052] 在一种实施方式中,还包括步骤S5:

[0053] S5、当所述哈希值的比对结果不满足所述冗余条件时,所述待测块为非冗余块,将所述非冗余块保存至图像存储器;其中,所述冗余条件为:

[0054] 所述待测块的哈希值和与所述待测块对应的原始块的哈希值之间的汉明距离小于预设距离阈值。

[0055] 示例性的,假设预设距离阈值为5,待测图像包括n个待测块,原始图像包括n个原始块,待测块和原始块一一对应,假设第i个待测块与对应的原始块的哈希值的汉明距离为4,小于预设距离阈值5,满足冗余条件,则说明第i个待测块和对应的原始块相似,第i个待测块为冗余块,删除第i个待测块;假设第j个待测块与对应的原始块的哈希值的汉明距离为10,大于预设距离阈值5,不满足冗余条件,说明第j个待测块与对应的原始块不相似,第j个待测块为非冗余块,此时将第j个待测块保存至图像存储器。通过对待测块的冗余检测,实现了冗余块的删除,非冗余块的保存,优化了图像的数据存储。

[0056] 在一种实施方式中,所述原始图像通过以下步骤获取:

[0057] S11、提取所述待测图像的标识信息;

[0058] S12、从图像存储器中选取标识信息与所述待测图像的标识信息相同的图像作为原始图像。

[0059] 具体地,由于待测图像为原始图像经图像修改得到的,因此,当原始图像中存在标识信息时,待测图像必定带有相同的标识信息。为了能够从图像存储器中准确快速地获取到与待测图像对应的原始图像,通过利用从待测图像中提取的标识信息查询图像存储器,从存储器中找到具备相同标识信息的图像作为原始图像。

[0060] 在一种实施方式中,在所述原始图像存入所述图像存储器之前,还包括步骤S01~S03:

[0061] S01、基于预设的图像分块规则,将所述原始图像分块得到若干原始块;

[0062] S02、获取原始标识信息;

[0063] S03、将所述原始标识信息添加到每一所述原始块中。

[0064] 示例性的,假设标识信息为鲁棒数字水印,原始标识信息为隐写水印元数据“User ID@DeviceID@San Jose,CA,US 1:23:45am 01/01/2020 3024x4032”。考虑到分块的复杂度

以及计算开销等因素,在原始图像生成时,将原始图像划分为若干固定大小的块,由于傅里叶变换的共轭对称特性,频谱在频域中是中心对称的,因此将元数据编码为镜像位图图像,随后将元数据图像和原始块转换到特定的特征空间,如余弦变换域、傅里叶变换域或小波变换域,然后将这些信号添加到特定的特征空间中,最后将结果转换回原始时域,得到了一份带有隐写水印的原始图像。进一步地,为了防止恶意软件伪造水印信息以伪造图像,在生成镜像位图图像之前对元数据进行加密,在提取水印后再对其进行解密。加密方案可以是凯撒加密方案,通过将元数据的每个字符替换为另一个可打印字符来隐藏纯文本图像描述信息。这种加密方案的开销可以忽略不计,因为只添加了几个shuffle指令,值得说明的是,加密方案也可以是其他加密方式,在此不作限定。

[0065] 优选的,原始图像的分块以及标识信息的嵌入过程与图像的生成相结合,例如将该功能添加到数码相机的固件中或者特定应用程序中,以使在原始图像生成时即可进行分块以及标识信息的添加。

[0066] 在一种实施方式中,所述原始标识信息通过以下方式生成:获取所述原始图像的描述信息,根据所述描述信息生成原始标识信息;其中,所述描述信息包括与所述原始图像生成相关的用户、设备、应用程序、日期、位置或分辨率中的至少一种。

[0067] 示例性的,原始标识信息为隐写水印,由诸如与图像生成相关的用户、设备/应用程序、日期、位置、分辨率等描述信息组成的位图图像,假设隐写水印元数据为“User ID@Device ID@San Jose,CA,US 1:23:45am 01/01/2020 3024x4032”,包括的描述信息有用户ID、设备ID、位置(圣何塞,加利福尼亚州,美国)、日期(2020年1月1日凌晨1点23分45秒)、分辨率(3024x4032)。

[0068] 在一种实施方式中,在步骤S4中的所述删除所述冗余块之前,还包括S41~S42:

[0069] S41、提取所述冗余块的图像信息,提取与所述冗余块对应的原始块的图像信息;

[0070] S42、根据所述冗余块的图像信息、与所述冗余块对应的原始块的图像信息,得到所述冗余块的图像操作信息,并将所述图像操作信息存储于图像描述存储器中;

[0071] 步骤S5中的所述将所述非冗余块保存至图像存储器,具体包括:

[0072] 将所述非冗余块附近的图像内容添加到所述非冗余块中,将添加所述图像内容后的所述非冗余块保存至所述图像存储器。

[0073] 示例性的,参见图2,待测图像B为原始图像A经过裁剪得到的,A由原始块A1~A24组成,经过图像对齐,B由待测块B1~B24组成,经过哈希值的比对,得到冗余块为B6、B7、B10、B11、B14、B15、B18、B19,非冗余块为B1、B2、B3、B4、B5、B8、B8、B12、B13、B16、B17、B20、B21、B22、B23、B24、B25,结合原始块进行计算,得到A6→B6、A7→B7、A10→B10、A11→B11、A14→B14、A15→B15、A18→B18、A19→B19的图像操作信息,由于B由A裁剪得到,因此,这些冗余块与对应的原始块完全相同,图像操作信息为简单的替换操作,图像操作信息被存储于图像描述存储器中,可通过调用对应的图像操作信息恢复消除的图像文件。为了便于后续的图像恢复,在保存非冗余块时,会将非冗余块附近的冗余内容一并进行保存,以帮助在图像恢复过程中将块重新组装成相应的图像。

[0074] 值得说明的是,添加到非冗余块的冗余内容的数量可根据实际需求进行设定;原始图像的修改方式也不局限于裁剪,还可以是缩放、旋转、添加滤镜、添加文本或更改颜色等,在此不作限定,此时的图像操作信息就不是简单的替换操作,图像操作信息可以划分为

块级描述信息和文件级描述信息,示例性的,假设冗余块x的明度为10,与该冗余块对应的原始块y的明度为50,则对应的图像操作信息就包括“将原始块y的明度减少40得到冗余块x”,该图像操作信息为块级描述信息;假设待测图像包括三个待测块a1、b1、c1,明度分别为1、2、3,待测图像的所有待测块都为冗余块(即a1、b1、c1都为冗余块),原始图像包括三个原始块a2、b2、c2,明度分别为4、5、6,a1和a2对应,b1和b2对应,c1和c2对应,那么经过计算可知,将原始块a2的明度减少3得到冗余块a1,将原始块b2的明度减少3得到冗余块b1,将原始块c2的明度减少3得到冗余块c1,即可推断出将原始图像的明度减少3可得到待测图像,对应的图像操作信息包括“将原始图像的明度减少3得到待测图像”,该图像操作信息为文件级描述信息。

[0075] 在一种实施方式中,在步骤S4中的所述删除所述冗余块之后,还包括步骤S6~S7:

[0076] S6、响应于图像恢复指令,根据所述图像操作信息和所述原始图像生成冗余图像;

[0077] S7、根据所述冗余图像和所述非冗余块,恢复所述待测图像。

[0078] 具体地,图像恢复是指将感知上等价的图像恢复到在图像重复数据删除阶段被冗余副本替换的图像的过程。根据图像存储器和图像描述存储器中的信息来部署相应的操作以恢复图像,通常包括三个步骤:转换、拼接和增强。转换阶段主要采取一系列操作对块进行恢复,如果图像在文件级别进行了重复数据删除,则可以绕过该阶段直接进行拼接。在拼接过程中,通过根据每个非冗余块边界周围额外的冗余内容,可以快速拼接所有块以生成候选图像,最后,增强阶段通过使用超分辨率恢复模型进一步优化图像,提高候选图像的质量,特别是块的边界,图像质量损失的原因来自于嵌入的隐写水印,但是该操作带来的质量损失在感知上并不明显,且通过采用超分辨率优化方法,生成最佳插值以降低均方误差,恢复后的图像的PSNR值(峰值信噪比)与嵌入隐写水印后的图像几乎相同。

[0079] 与现有技术相比,本发明实施例实现了块级的感知冗余内容的删除,降低了直接以待测图像作为冗余检测对象来检测冗余的计算量和计算延迟,同时也能针对整体相似性较小的两个图像中的冗余部分进行有效删除,减少存储空间的浪费,优化了图像的数据存储。

[0080] 为了更好地说明本发明实施例的图像处理方法的优点,以下对实际应用例子的图像处理结果进行比对。

[0081] 在一实际应用例子中,图像数据库中的原始图像有1.3GB,共有362.5GB副本的数据集,其中重复的副本共有114.2GB,通过基于相似度的重复数据消除方法(similarity-based deduplication, SIM-dedup)过滤掉可区分副本后剩余22.8GB,而本发明实施例的图像数据处理方法可以将数据进一步减少到6.7GB,相比于SIM-dedup节省了70.6%的存储空间,冗余图像放大因子由SIM-dedup的17.5 ($22.8 \div 1.3 = 17.5$)减少到5.2 ($6.7 \div 1.3 = 5.2$),其中,冗余图像放大因子定义为实际图像数据大小除以原始图像数据大小,可参见图3所示的图像存储空间要求柱形图,纵轴为存储空间大小,横轴从左到右分别为数据集的所有副本、重复的副本、经SIM-dedup后的副本、原始图像、本发明实施例的图像处理方法处理后的副本,由图3可知,本发明实施例能够对图像的冗余部分进行有效删除,减少存储空间的浪费,对存储空间要求较低。

[0082] 在执行时间方面, SIM-dedup将大部分时间用于提取基于VGG16的高维特征向量,而WM-dedup在特征提取方面(水印提取)所花的时间少得多。本方案将大部分时间都花在恢

复图像上,利用感知上等效的块重组图像,并在任意两个相邻块之间的所有边界上应用超分辨率优化以尽可能恢复去重的图像。SIM-dedup恢复图像的操作非常简单,延迟主要来自于选择和加载近乎重复的图像,但是图像还原效果较差。具体来说,本方法提取特征的速度是SIM-dedup的5.7倍,重复数据删除的速度是SIM-dedup的4.5倍,虽然WM-dedup使用的块级重复数据删除方式会增加查询次数,但是由于查询操作所需的时间在微秒级,因此不会对系统性能造成明显的影响。虽然本方法在图像恢复阶段花费的时间比SIM-dedup多得多,但是本方法的总体执行时间仍然比SIM-dedup缩短了接近60%,详细的各阶段执行时间比对数据请参见下表。

[0083]		特征提取	图像哈希	查询(寻找原始图像)	去重	恢复
	SIM-dedup	482ms	4ms	25us/文件	502ms	8ms
	本方法	84ms	4ms	12us/块	112ms	195ms

[0084] 参见图4,本发实施例还提供一种图像数据处理装置,包括:

[0085] 图像获取模块11,用于获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;

[0086] 图像分块模块12,用于基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;

[0087] 图像比对模块13,用于分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;

[0088] 图像删除模块14,用于当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。

[0089] 值得说明的是,具体的所述图像数据处理装置的工作过程可参考上述实施例中所述图像数据处理方法的工作过程,在此不再赘述。

[0090] 与现有技术相比,本发明实施例提供的装置能够基于图像对齐算法,将获取的待测图像和原始图像进行对齐;结合原始图像的分块情况,对待测图像进行分块处理得到待测块,并建立其原始图像的原始块和待测图像的待测块之间的一一对应的映射关系;基于映射关系进行待测块和原始块之间的哈希值比对,进而根据哈希值的比对结果来确定冗余块,以用于删除冗余块,实现了块级的感知冗余内容的删除,降低了直接以待测图像作为冗余检测对象来检测冗余的计算量和计算延迟,同时也能针对整体相似性较小的两个图像中的冗余部分进行有效删除,减少存储空间的浪费,优化了图像的数据存储。

[0091] 本发明实施例还提供一种图像数据处理设备,包括处理器、存储器以及存储在所述存储器中且被配置为由所述处理器执行的计算机程序,所述处理器执行所述计算机程序时实现如上述图像数据处理方法实施例中的步骤,例如图1中所述的步骤S1~S4;或者,所述处理器执行所述计算机程序时实现上述各装置实施例中各模块的功能,例如参考因素获取模块。

[0092] 示例性的,所述计算机程序可以被分割成一个或多个模块,所述一个或者多个模块被存储在所述存储器中,并由所述处理器执行,以完成本发明。所述一个或多个模块可以是能够完成特定功能的一系列计算机程序指令段,该指令段用于描述所述计算机程序在所述图像数据处理设备中的执行过程。例如,所述计算机程序可以被分割成多个模块,各模块具体功能如下:

[0093] 图像获取模块11,用于获取待测图像和原始图像;其中,所述原始图像由若干原始块组成;

[0094] 图像分块模块12,用于基于图像对齐算法,根据所述原始图像对所述待测图像进行分块得到若干待测块,并建立所述待测块和所述原始块的映射关系;

[0095] 图像比对模块13,用于分别计算所述待测块和与所述待测块对应的原始块的哈希值,并进行哈希值比对;

[0096] 图像删除模块14,用于当所述哈希值的比对结果满足预设的冗余条件时,所述待测块为冗余块,删除所述冗余块。

[0097] 各个模块具体的工作过程可参考上述实施例所述的图像数据处理装置的工作过程,在此不再赘述。

[0098] 所述图像数据处理设备可以是桌上型计算机、笔记本、掌上电脑及云端服务器等计算设备。所述图像数据处理设备可包括,但不限于,处理器、存储器。本领域技术人员可以理解,所述示意图仅仅是图像数据处理设备的示例,并不构成对图像数据处理设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述图像数据处理设备还可以包括输入输出设备、网络接入设备、总线等。

[0099] 所述处理器可以是中央处理单元(Central Processing Unit,CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field-Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等,所述处理器是所述图像数据处理设备的控制中心,利用各种接口和线路连接整个图像数据处理设备的各个部分。

[0100] 所述存储器可用于存储所述计算机程序和/或模块,所述处理器通过运行或执行存储在所述存储器内的计算机程序和/或模块,以及调用存储在存储器内的数据,实现所述图像数据处理设备的各种功能。所述存储器可主要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、至少一个功能所需的应用程序等;存储数据区可存储根据手机的使用所创建的数据等。此外,存储器可以包括高速随机存取存储器,还可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card,SMC),安全数字(Secure Digital,SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件、或其他易失性固态存储器件。

[0101] 其中,所述图像数据处理设备集成的模块如果以软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明实现上述实施例方法中的全部或部分流程,也可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一计算机可读存储介质中,该计算机程序在被处理器执行时,可实现上述各个方法实施例的步骤。其中,所述计算机程序包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。

[0102] 以上所述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,这些改进和润饰也视为本发明的保护范围。

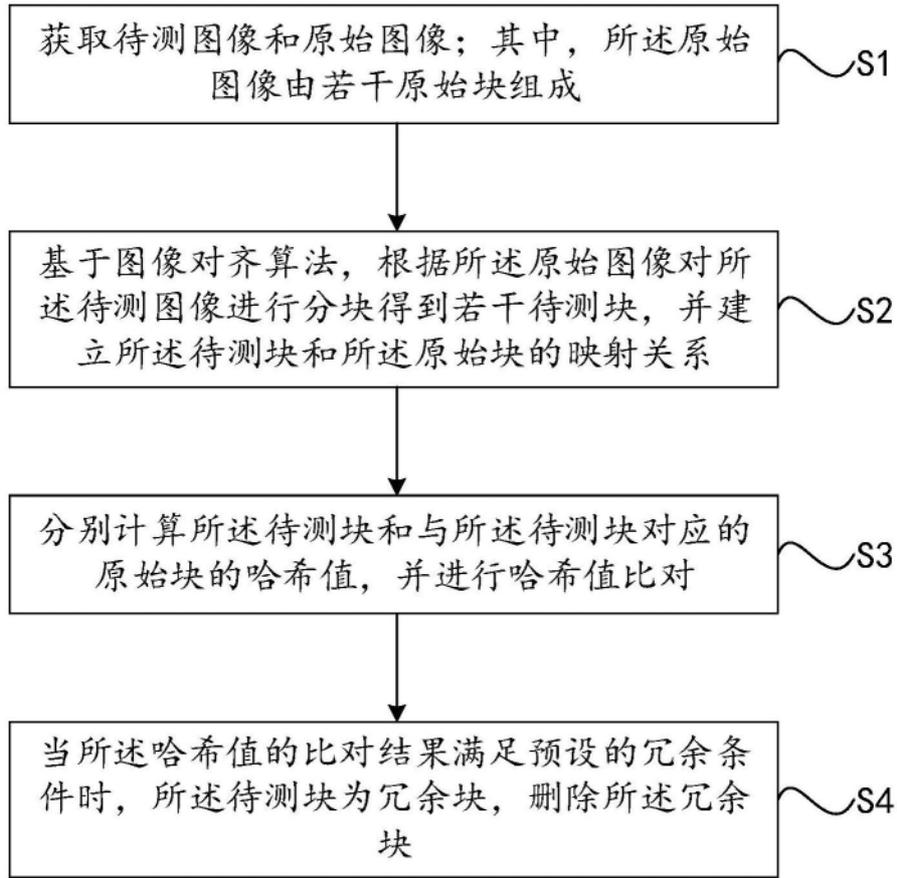


图1



图2

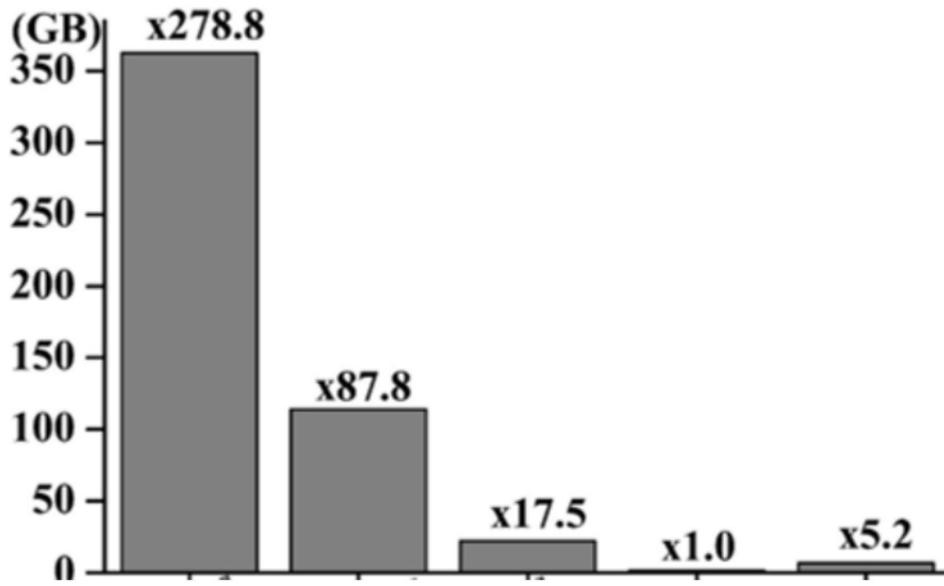


图3

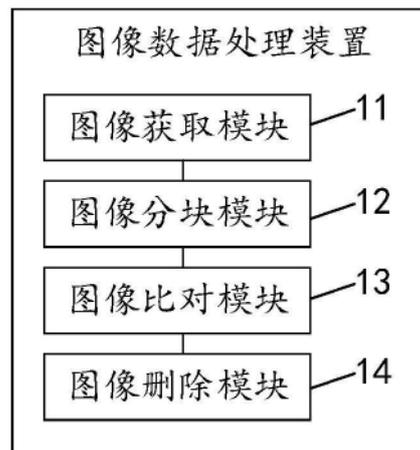


图4