(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
2016 年 9 月 15 日 (15.09.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/141803 A1

(51) 国际专利分类号:
*G06K 9/00* (2006.01)

(21) 国际申请号: PCT/CN2016/074240

(22) 国际申请日: 2016 年 2 月 22 日 (22.02.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510101155.9 2015 年 3 月 6 日 (06.03.2015) CN

(71) 申请人: **华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD)** [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **南洋理工大学 (NANYANG TECHNOLOGICAL UNIVERSITY)** [SG/SG]; 新加坡南洋道 50 号, Singapore 639798 (SG)。

(72) 发明人: **余浩 (YU, Hao)**; 新加坡南洋道 50 号, singapore 639798 (SG)。 **王雨豪 (WANG, Yuhao)**; 新加坡南洋道 50 号, Singapore 639798 (SG)。 **倪磊滨 (NI, Leibin)**; 新加坡南洋道 50 号, Singapore 639798 (SG)。 **杨伟 (YANG, Wei)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **赵俊峰 (ZHAO, Junfeng)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 **肖世海 (XIAO, Shihai)**; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

[见续页]

(54) **Title**: IMAGE RECOGNITION ACCELERATOR, TERMINAL DEVICE AND IMAGE RECOGNITION METHOD

(54) **发明名称** : 图像识别加速器、终端设备及图像识别方法

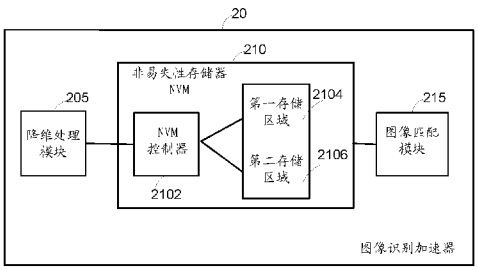


图 3

20 Image recognition accelerator
205 Dimension reduction processing module
210 Non-volatile memory (NVM)
215 Image matching module
2102 NVM controller
2104 First storage region
2106 Second storage region

(57) **Abstract**: An image recognition accelerator (20), a terminal device and an image recognition method. The image recognition accelerator (20) comprises a dimension reduction processing module (205), an NVM (210) and an image matching module (215). In a process where the image recognition accelerator (20) performs image recognition, firstly, the dimension reduction processing module (205) reduces the dimension of first image data according to a set dimension reduction parameter γ. The NVM (210) writes a low ω bit of each numerical value in the first image data after the dimension reduction into a first storage region (2104) of the NVM (210) according to a set first current I, and writes a high N-ω bit of each numerical value in the first image data after the dimension reduction into a second storage region (2106) of the NVM (210) according to a set second current, wherein the first current is less than the second current. Thus, the matching module (215) can determine whether an image library stored in the NVM contains image data matching the first image data after the dimension reduction. The image recognition accelerator (20) can ensure the accuracy of image recognition on the basis of reducing the system power consumption of a terminal device.

(57) **摘要**:

[见续页]

WO 2016/141803 A1

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))。

一种图像识别加速器(20)、终端设备及图像识别方法。图像识别加速器(20)包括了降维处理模块(205)、NVM(210)以及图像匹配模块(215)。在图像识别加速器(20)进行图像识别的过程中，先由降维处理模块(205)根据设置的降维参数γ降低第一图像数据的维度。NVM(210)将降维后的第一图像数据中的各个数值的低ω位按照设置的第一电流 I 写入 NVM(210)中的第一存储区域(2104)，并将降维后的第一图像数据中的各个数值的高 N-ω 位按照设置的第二电流写入 NVM (210) 中的第二存储区域(2106)。其中，第一电流小于第二电流。从而，匹配模块(215)可以确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。图像识别加速器(20)能够在降低终端设备的系统功耗的基础上保证图像识别的准确性。

# 图像识别加速器、终端设备及图像识别方法

## 技术领域

**[0001]** 本发明涉及计算机技术领域，尤其涉及一种图像识别加速器、终端设备及图像识别方法。

## 背景技术

**[0002]** 图像识别技术是人工智能的一个重要领域。图像识别是指利用计算机对图像进行处理和分析，以识别各种不同目标和对像的技术。近年来，随着社交网络的普及，在移动设备中进行实时图像数据分析的需求逐渐增强。然而，由于实现图像数据分析会消耗较多的系统资源，因此，移动设备有限的电池寿命限制了图像数据分析在移动设备上的应用。

**[0003]** 为了降低图像数据分析过程中的系统功耗，现有技术中的一种图像数据处理方法是通过降低图像数据写入静态随机存储器（Static Random-Access Memory, SRAM）中的写入电流的方式来降低系统功耗。然而，随着写入电流的降低，SRAM 中存储的数据的错误率也随之上升。为了恢复错误，还需要通过解凸优化（convex optimization）处理等方式对存储的图像数据进行恢复，从而能够基于恢复后的图像数据进行图像识别。在这种方式中，虽然写入数据时系统功耗有所减少，然而，图像恢复过程中的 CPU 的计算复杂度高，也比较浪费系统资源。并且，为了保持存储于 SRAM 中的数据，SRAM 需要保持通电状态，因此，SRAM 还存在静态功耗，上述图像数据处理方式也并不能完全消除 SRAM 保持数据时所需的静态功耗。因此，总体上说，采用现有的图像数据处理方式处理图像数据时，系统功耗依然较大。

## 发明内容

**[0004]** 本发明实施例中提供的一种图像识别加速器、终端设备及图像识别方法，能够在降低终端设备的系统功耗的基础上保证图像识别的准确性。

**[0005]** 第一方面，本发明实施例提供了一种应用于终端设备中用于识别图像的图像识别加速器，包括:

降维处理模块，用于根据设置的降维参数γ降低第一图像数据的维度，其中，降维后的第一图像数据包括多个数值;

非易失性内存NVM，用于将降维后的第一图像数据的每一个数值的低ω位按照设置

的第一电流I存储于所述NVM的第一存储区域，将降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，ω为设置的宽度参数，所述第一电流I小于所述第二电流$I_s$，所述降维参数γ、宽度参数ω以及第一电流I根据所述终端设备的系统功耗和设置的第一图像识别成功率获得；

图像匹配模块，用于确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

**[0006]** 结合第一方面，在第一方面的第一种可能的实现方式中，所述图像识别加速器还包括：参数调整模块，用于如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数 γ、宽度参数 ω 以及第一电流 I，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

所述降维处理模块，还用于根据调整后的降维参数γ'降低第二图像数据的维度；

所述非易失性内存NVM，还用于将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，ω'为调整后的宽度参数，所述 I'小于所述$I_s$；

所述图像匹配模块，还用于确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

**[0007]** 结合第一方面或第一方面的第一种实现方式，在第一方面的第二种可能的实现方式中，所述降维处理模块具体用于：

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的值根据设置的降维参数γ确定，γ=n/m。

**[0008]** 结合第一方面的第一种或第二种可能的实现方式，在第一方面的第三种可能的实现方式中，所述参数调整模块具体用于：

如果统计的图像识别成功率与所述第二图像识别成功率之间的差值大于预设阈值，则分别调整降维参数γ、所述宽度参数ω或第一电流I的取值以降低系统功耗E，并分别获得调整后的图像识别成功率，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比；

确定在调整后的图像识别成功率与所述第二图像识别成功率之间的差值的绝对值不大于所述预设阈值时所述终端设备的最小功耗E';

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数γ、所述宽度参数ω以及第一电流I分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

**[0009]** 第二方面，本发明实施例提供了一种终端设备，所述终端设备包括CPU和图像识别加速器，其中，所述CPU，用于向所述图像识别加速器发送待识别的第一图像数据;

所述图像识别加速器，用于根据设置的降维参数γ降低所述第一图像数据的维度，其中，降维后的第一图像数据包括多个数值;

将所述降维后的第一图像数据的每一个数值的低ω位按照设置的第一电流I存储于NVM的第一存储区域，将所述降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，ω为设置的宽度参数，所述I小于所述$I_s$，所述降维参数γ、宽度参数ω以及第一电流I根据所述终端设备的系统功耗和设置的第一图像识别成功率获得;

确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

**[0010]** 结合第二方面，在第二方面的第一种可能的实现方式中，所述图像识别加速器，还用于如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，其中，所述第二图像识别成功率与所述第一图像识别成功率不同;

所述CPU，还用于向所述图像识别加速器发送第二图像数据;

所述图像识别加速器，还用于:

根据调整后的降维参数γ'降低第二图像数据的维度;

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，ω'为调整后的宽度参数，所述 I'小于所述$I_s$;

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

**[0011]** 结合第二方面，在第二方面的第二种可能的实现方式中，所述 CPU，还用于统

计在预设的统计期间内所述图像识别加速器输出的匹配结果，获取统计的图像识别成功率；确定所述统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值；

所述图像识别加速器，还用于根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

所述CPU，还用于向所述图像识别加速器发送第二图像数据；

所述图像识别加速器，还用于根据调整后的降维参数γ'降低第二图像数据的维度；

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述ω'为调整后的宽度参数，所述 I'小于所述$I_s$；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

**[0012]** 结合第二方面，在第二方面的第三种可能的实现方式中，所述 CPU 还用于：

统计在预设的统计期间内所述图像识别加速器输出的匹配结果，获取所述统计的图像识别成功率；

如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

向所述图像识别加速器发送第二图像数据；

所述图像识别加速器，还用于：

根据调整后的降维参数γ'降低第二图像数据的维度；

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，ω'为调整后的宽度参数，所述 I'小于所述$I_s$；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

**[0013]** 结合第二方面或第二方面的第一种至第三种任意一种可能的实现方式中，在第二方面的第四种可能的实现方式中，所述图像识别加速器具体用于:

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的值根据设置的降维参数γ确定，γ=n/m。

**[0014]** 结合第二方面的第一种或第二种可能的实现方式中，在第二方面的第五种可能的实现方式中，所述图像识别加速器具体用于:

分别调整降维参数γ、所述宽度参数ω或第一电流I的取值以降低所述终端设备的系统功耗E，并分别获得调整后的图像识别成功率，其中，所述E的值与γ((N-ω)* $I_s^2$ +ω*I)的值成正比;

确定在调整后的图像识别成功率与所述第二图像识别成功率之间的差值的绝对值不大于所述预设阈值时所述终端设备的最小功耗E';

选择在满足所述最小功耗 E'时获得最大图像识别成功率的降维参数 γ、所述宽度参数 ω以及第一电流 I 分别作为所述调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。

**[0015]** 结合第二方面的第三种可能的实现方式中，在第二方面的第六种可能的实现方式中，所述 CPU 具体用于:

分别调整降维参数γ、所述宽度参数ω或第一电流I的取值以降低所述终端设备的系统功耗E，并分别获得调整后的图像识别成功率，其中，所述E的值与γ((N-ω)* $I_s^2$ +ω*I)的值成正比;

确定在调整后的图像识别成功率与所述第二图像识别成功率之间的差值的绝对值不大于所述预设阈值时所述终端设备的最小功耗E';

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数γ、所述宽度参数ω以及第一电流I分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

**[0016]** 第三方面，本发明实施例提供了一种应用于终端设备的图像识别方法，所述方法由所述终端设备中的图像识别加速器执行，所述方法包括:

根据设置的降维参数γ降低第一图像数据的维度，其中，降维后的第一图像数据包括多个数值;

将所述降维后的第一图像数据的每一个数值的低ω位按照设置的第一电流I存储于所述图像识别加速器中的NVM的第一存储区域，将所述降维后的第一图像数据的每一个

数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，ω为设置的宽度参数，所述I小于所述$I_s$，所述降维参数γ、宽度参数ω以及第一电流I根据所述终端设备的系统功耗和设置的第一图像识别成功率获得;

确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

**[0017]** 结合第三方面，在第三方面第一种可能的实现方式中，所述方法还包括:

确定统计的图像识别成功率与设置的第二图像识别成功率的差值的绝对值大于预设阈值;

根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，其中，所述第二图像识别成功率与所述第一图像识别成功率不同;

根据调整后的降维参数γ'降低第二图像数据的维度;

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，ω'为调整后的宽度参数，所述 I'小于所述$I_s$;

确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

**[0018]** 结合第三方面或第三方面的第一种可能的实现方式，在第三方面的第二种可能的实现方式中，所述根据设置的降维参数 γ 降低第一图像数据的维度包括:

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的值根据设置的降维参数γ确定，γ=n/m。

**[0019]** 结合第三方面或第三方面的第一种至第二种中任意一种可能的实现方式，在第三方面的第三种可能的实现方式中，所述根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数 γ、宽度参数 ω 以及第一电流 I，包括:

分别调整降维参数γ、所述宽度参数ω或第一电流I的取值以降低所述终端设备的系统功耗E，并分别获得调整后的图像识别成功率，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比;

确定在调整后的图像识别成功率与所述第二图像识别成功率之间的差值的绝对值不大于所述预设阈值时所述终端设备的最小功耗E';

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数γ、所述宽度参数ω以及第一电流I分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

**[0020]** 第四方面，本发明实施例提供了一种计算机程序产品，包括存储了程序代码的计算机可读存储介质，所述程序代码包括的指令用于执行前述第三方面中所述的方法。

**[0021]** 第五方面，本申请提供了又一种应用于终端设备中用于识别图像的图像识别加速器。该图像识别加速器包括降维处理模块、非易失性内存 NVM 以及图像匹配模块。所述降维处理模块用于接收降维参数 γ，根据接收的降维参数 γ 降低第一图像数据的维度，其中，降维后的第一图像数据包括多个数值，所述降维参数 γ 是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的。所述非易失性内存 NVM 用于接收宽度参数 ω 和第一电流 I，并根据接收的宽度参数 ω 获得存储位数 S，然后，将降维后的第一图像数据的每一个数值的低 S 位按照设置的第一电流 I 存储于所述 NVM 的第一存储区域，将降维后的第一图像数据的每一个数值的高(N-S)位按照设置的第二电流 $I_s$ 存储于所述 NVM 的第二存储区域.其中，N 为每一个数值所占的比特位，所述第一电流 I 小于所述第二电流 $I_s$，所述宽度参数 ω 以及所述第一电流 I 是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的。所述图像匹配模块，用于确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

**[0022]** 结合第五方面，在一种可能的实现方式中，所述图像识别加速器还包括参数调整模块。所述参数调整模块用于根据设置的所述第一图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数、宽度参数以及第一电流的值，以获得调整后的降维参数 γ、宽度参数 ω 以及第一电流 I，并将调整后的所述降维参数 γ 发送给所述降维处理模块，将调整后的所述宽度参数 ω 以及调整后的所述第一电流 I 发送给所述 NVM。

**[0023]** 结合上述第五方面以及可能的实现方式，在又一种可能的实现方式中，所述参数调整模块具体用于：分别调整所述降维参数、所述宽度参数或所述第一电流的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，每个调整后的图像识别成功率对应于每个调整后的系统功耗；确定调整后的每个图像识别成功率与所述第一图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗；选择在满足所述最小系统功耗时获得最大图像识别成功率的降维参数、宽度参数以及第一电流的值分别作为调整后的所述

降维参数 γ、所述宽度参数 ω 以及所述第一电流 I，并将调整后的所述降维参数 γ 发送给所述降维处理模块，将调整后的所述宽度参数 ω 以及所述第一电流 I 发送给所述 NVM。

**[0024]** 结合上述第五方面以及可能的实现方式，在又一种可能的实现方式中，所述降维处理模块具体用于：根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为 k 行*m 列的矩阵，所述二进制矩阵为 m 行*n 列的矩阵，所述降维后的第一图像数据为 k 行*n 列的矩阵，k、m 和 n 为正整数，m 的值大于 n，n 的值根据设置的降维参数 γ 确定，γ=n/m。

**[0025]** 本发明实施例提供的一种应用于终端设备中用于进行图像识别的图像识别加速器包括了降维处理模块、NVM 以及图像匹配模块。在所述图像识别加速器对第一图像数据进行识别的过程中，先由降维处理模块根据设置的降维参数 γ 降低第一图像数据的维度。NVM 可以将降维后的第一图像数据中的各个数值的低 ω 位按照设置的第一电流 I 写入 NVM 中的第一存储区域，并将降维后的第一图像数据中的各个数值的高 N-ω 位按照设置的第二电流 $I_s$ 写入 NVM 中的第二存储区域。其中，第一电流小于第二电流。从而，匹配模块可以确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据，以获得第一图像数据的图像识别结果。由于设置的降维参数 γ、宽度参数 ω 以及第一电流 I 均是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得，因此能够保证存储于第一存储区域的数值中的低位部分在存储过程中出现的错误对第一图像数据的识别成功率的影响较小。本发明实施例提供的图像识别加速器能够在降低终端设备的系统功耗的基础上保证图像识别的准确性，并且能够提高图像数据的识别速度。

## 附图说明

**[0026]** 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中的附图。

**[0027]** 图 1 为本发明实施例提供的一种终端设备的结构示意图；

**[0028]** 图 2 为本发明实施例提供的另一种终端设备的结构示意图；

**[0029]** 图 3 为本发明实施例提供的一种图像识别加速器的结构示意图；

**[0030]** 图 4 为本发明实施例提供的一种图像识别方法的流程图；

**[0031]** 图 5 为本发明实施例提供的一种降维处理模块的结构示意图；

**[0032]** 图 6 为本发明实施例提供的一种 NVM 的硬件结构示意图；

**[0033]** 图 7 为本发明实施例提供的又一种图像识别加速器的结构示意图；

**[0034]** 图 8 为本发明实施例提供的又一种图像识别方法的流程图；

**[0035]** 图 9 为本发明实施例提供的一种参数调整方法流程图；

**[0036]** 图 10 (a)和图 10 (b)为本发明实施例提供的参数调整过程中记录参数的示意图；

**[0037]** 图 11 为本发明实施例提供的又一种终端设备的结构示意图；

**[0038]** 图 12 为本发明实施例提供的又一种终端设备结构示意图；

**[0039]** 图 13 为本发明实施例提供的一种图像识别方法的信令图。

## 具体实施方式

**[0040]** 为了使本技术领域的人员更好地理解本发明方案，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的实施例，而不是全部的实施例。

**[0041]** 本发明实施例提出了一种图像识别加速器，能够在降低系统功耗的基础上保证图像识别的准确性。图 1 为本发明实施例提供的一种终端设备的结构示意图。在图 1 所述的终端设备 100 中，中央处理器(Central Processing Unit，CPU)10 与图像识别加速器 20 通过总线 15 直接交换数据。总线 15 可以为 PCI、PCIE 总线或图像加速接口 AGP 总线等系统总线，在本发明实施例中不对总线 15 的类型进行限定。需要说明的是，图 1 所示的终端设备 100 可以是计算机、手机、移动终端等终端设备，在此不做限定，只要是需要实现图像识别的终端设备即可。

**[0042]** 如图 1 所示， CPU 10 是终端设备 100 的运算核心（Core）和控制核心（Control Unit）。CPU 10 可以是一块超大规模的集成电路。在 CPU 10 中安装有操作系统和其他软件程序，从而 CPU 10 能够实现对内存、缓存等存储空间的访问。可以理解的是，在本发明实施例中，CPU 10 仅仅是处理器的一个示例。除了 CPU 10 外，处理器还可以是其他特定集成电路（Application Specific Integrated Circuit, ASIC），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。

**[0043]** 图像识别加速器 20 是硬件加速器（Hardware accelerator）中的一种。在本发明实施例中，图像识别加速器 20 是基于非易失内存（Non-Volatile Memory, NVM）的硬件加速器。硬件加速技术是利用硬件模块来替代软件算法，从而可以充分利用硬件所固有的快速特性，以提高计算机系统的处理速度。在传统的图像数据处理方法中，内存仅仅用于存储图像

数据，所有的图像数据处理、分析工作均由 CPU 完成，因此，CPU 的处理速度以及内存的传输带宽成为图像识别技术发展的瓶颈。在本发明实施例中，通过在内存中增加简单的逻辑处理电路，以实现由专用的图像识别加速器来实现图像数据处理工作。在图 1 所示的终端设备 100 中，CPU 10 只需要向图像识别加速器 20 发送待识别的图像数据并接受图像识别加速器 20 获得的识别结果，从而减少了 CPU 10 的负担，提高了终端设备识别图像的速度。并且，由于图 1 所示的系统结构减少了在 CPU 10 和图像识别加速器 20 的之间传输的数据量，从而可以解决因内存的传输带宽限制图像识别的速度的问题。

**[0044]** 实际应用中，图像识别加速器 20 并不唯一与 CPU 10 进行数据交互。图 2 为本发明实施例提供的另一种终端设备的结构示意图。图 2 所示的终端设备 100 可以包括 CPU 10、图像识别加速器 20 以及图像数据采集器 30。CPU 10 和图像数据采集器 30 分别与图像识别加速器 20 连接。图像数据采集器 30，用于采集图像数据信息，并将采集的图像数据信息发送给图像识别加速器 20 进行图像识别。本领域人员可以知道，图像数据采集器 30 可以采集人、物的图像，在此不对图像信息进行具体的限定。图像数据采集器 30 在采集到图像信息后，可以将采集的图像信息转换为图像数据。实际应用中，图像数据采集器可以包括实现照相或摄像等功能的器件。例如，图像数据采集器可以为手机上的摄像头。图像识别加速器 20，用于对图像数据采集器 30 发送的图像数据信息与存储的图像数据信息进行识别，并将识别结果发送给 CPU 10。可以理解的是，在图 2 中所示的 CPU 10 以及图像识别加速器 20 的功能及实现方式的描述可以参见上述图 1 的描述，在此不再赘述。

**[0045]** 可以理解的是，上面描述的仅仅是本发明实施例中的终端设备 100 的两种示意性结构，说明了图像识别加速器 20 的两种应用场景。在又一种情形下，图像识别加速器 20 还可以接收 CPU 10 发送的图像数据信息进行图像识别后，将图像识别结果发送给其他器件或设备。在又一种情形下，图像识别加速器 20 还可以接收其他器件（例如图 2 中的图像数据采集器 30）发送的图像数据信息，并向该器件反馈图像识别结果。在此不对与图像识别加速器 20 通信的器件进行限定。下面将对本发明实施例提供的图像识别加速器 20 的具体结构及操作过程进行详细描述。

**[0046]** 图 3 为本发明实施例提供的一种图像识别加速器 20 的结构示意图，在图 3 中，对图像识别加速器 20 的结构进行了较为详细的图示。如图 3 所示，在本发明实施例中，图像识别加速器 20 可以包括降维处理模块 205、非易失性存储器 NVM 210、以及图像匹配模块 215。需要说明的是，降维处理模块 205 以及图像匹配模块 215 均可以是逻辑电路的形式

存在，也可以是以集成电路的形式存在。实际应用中，图像识别加速器 20 可以是一种特定集成电路 ASIC（Application Specific Integrated Circuit）或一种单板。在本发明实施例中，不对图像识别加速器 20 的具体存在形式进行限定。为了清楚的描述图像识别加速器 20 中各个器件的工作原理，下面将结合图 4 所示的图像识别方法的流程图对图 3 所示的图像识别加速器 20 中的各个器件的结构及工作过程进行详细的介绍。在下面的实施例中，以图像识别加速器 20 处理第一图像数据为例进行描述。

**[0047]** 降维处理模块 205，用于根据设置的降维参数 γ 降低图像数据的维度。具体的，如图 4 所示，在步骤 400 中，降维处理模块 205 可以根据设置的降维参数 γ 降低第一图像数据的维度。本领域技术人员可以知道，图像数据是用数值表示的各像素（pixel）的灰度值的集合。通常，图像数据是通过顺序地抽取图像的每一个像素的信息而获得的一个离散的阵列，该离散的阵列可以代表一副连续的图像。例如，第一图像数据可以表示为一个 k 行*m 列的矩阵，其中该矩阵中的每一个数值用于代表第一图像数据中的一个像素的灰度值。换一种表达方式，第一图像数据是用数值表示的第一图像的各像素的灰度值的集合。为了提高图像识别的速度，在本发明实施例中降维处理模块 205 可以采用基于稀疏表示的随机映射方式降低第一图像数据的维度。在本发明实施例中并不对降维处理模块 205 实现的压缩算法进行限定，只要是能够实现通过稀疏表示的随机映射算法均可。

**[0048]** 实际应用中，降维处理模块 205 可以采用矩阵乘法器来实现。具体的，降维处理模块 205 可以采用矩阵乘法器将第一图像数据与设置的低维的二进制矩阵相乘，从而实现降低第一图像数据的维度的目的。其中，二进制矩阵是指矩阵中的所有数值均为采用 0 或 1 表示的矩阵。采用二进制矩阵来实现降维的目的是为了减少降维过程中的计算复杂度。在本发明实施例中，设置的低维的二进制矩阵可以为伯努利矩阵，但本发明实施例不限定具体的二进制矩阵的形式，只要能够通过稀疏表示的方式实现降维目的的二进制矩阵即可。例如，第一图像数据为 k 行*m 列的矩阵 X，设置的二进制矩阵为 m 行*n 列的伯努利矩阵 Z，其中，k、m 和 n 均为正整数，且 m 大于 n。通过矩阵乘法器可以将第一图像数据与设置的伯努利矩阵 Z 相乘，从而可以获得一个 k 行*n 列矩阵 Y，矩阵 Y 即为降维后的第一图像数据。换一种表达方式，降低 X 矩阵的维度实际是为了降低 X 矩阵的列的数量。实际应用中，n 的值可以根据 m 的值以及设置的降维参数 γ 来确定，其中，降维参数 γ 为降维后的第一图像数据的维度与第一图像数据的维度的比值，即 γ=n/m，则 n=m*γ。降维参数 γ 也可以被称为是降维率。

**[0049]** 本领域人员可以知道，乘法器（multiplier）是一种用于完成两个互不相关的模拟

信号或数字信号相乘的作用的电子器件。乘法器可以将两个二进制数相乘。矩阵乘法器是由多个乘法器和加法器构成的用以实现矩阵与矩阵相乘的功能的器件。由于矩阵乘法器中不同列的乘法器和加法器的计算互不相关，能够实现并行运算，因此，可以通过增加或减少矩阵乘法器中乘法器及加法器的列数来调整矩阵的维数。为了描述方便，在本发明实施例中，将乘法器和加法器构成的用以实现矩阵乘法运算的电路简称为乘加器。

**[0050]** 在本发明实施例中，降维处理模块 205 可以通过减少矩阵乘法器中的部分列的乘加器来实现降低图像数据的维度的目的。具体的，可以通过关闭降维处理模块 205 中的部分列的乘加器的电源来减少部分列的乘加器。图 5 为本发明实施例提供的一种降维处理模块 205 的结构示意图。如图 5 所示，降维模块 205 包括 m 列乘加器，各列乘加器之间的运算互相独立。各列乘加器通过独立的开关控制该列乘加器是否工作。例如，开关 S1 用于控制第 1 列乘加器，开关 S2 用于控制第 2 列乘加器，依次类推，开关 Sm 用于控制第 m 列乘加器。本领域技术人员可以知道，开关可以通过场效应晶体管或开关电路来实现，例如，开关可以为结型场效应管(junction field effect transistor, JFET)和金属氧化物半导体场效应管(metal-oxide semiconductor FET, MOS-FET)，在此不对开关的实现方式进行限制。

**[0051]** 例如，在本发明实施例中，降维处理模块 205 可以接收 CPU 10 或图像数据采集器 30 发送的第一图像数据 X，其中，X 为 k 行*m 列的矩阵。假设设置的伯努利矩阵 Z 为 m 行*n 列的矩阵。降维处理模块 205 中可以设置 m 列乘加器。在一个周期内，可以将第一图像数据中的一个数值分别传输到矩阵乘法器的 m 列乘加器。矩阵乘加器中的各列乘加器可以分别将接收该数值与降维处理模块 205 中存储的伯努利矩阵 Z 中的一行数值中的一个数值进行乘法运算，并输出计算结果。换一种表达方式，每列乘加器一个周期内能够计算 X 矩阵中的一个数值与 Z 矩阵中的一个数值的计算结果，则在一个周期内，m 列乘加器能够获得 X 矩阵中的该数值与 Z 矩阵中的一行数值的计算结果。可以理解的是，根据这种方式，经过 m*k 个循环后，可以获得 X 矩阵中的 K 行数值与伯努利矩阵 Z 的运算结果。在降低第一图像数据的维度的过程中，为了实现对第一图像数据的降维操作，降维处理模块 205 可以根据设定的降维参数 γ 以及第一图像数据中的 m 的值获得 n 的值，并根据获得的 n 的值关闭矩阵乘法器中控制 m-n 列乘加器的开关。例如，如图 5 所示，可以关闭矩阵乘法器中控制第 n+1 列至第 m 列乘加器的开关，使矩阵乘法器中的第 n+1 列至第 m 列乘加器在运算的过程中不进行运算。根据上述方式，降维处理模块 205 能够实现 X 矩阵和 Z 矩阵的乘法运算，获得降维后的第一图像数据，其中，降维后的第一图像数据以 k 行*n 列的 Y 矩阵来表示。

**[0052]** 非易失性内存（Non-Volatile Memory, NVM）210，用于存储待识别的图像数据

以及预设的图像库中的图像数据。具体的，如图 4 所示，在步骤 410 中，NVM 210 可以将降维后的第一图像数据的每一个数值的低 ω 位按照设置的第一电流 I 存储于所述图像识别加速器 20 中的 NVM 210 的第一存储区域 2104，将降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流 $I_s$ 存储于所述 NVM 210 的第二存储区域 2106。其中，所述降维参数 γ、宽度参数 ω 以及第一电流 I 根据所述终端设备的系统功耗和设置的第一图像识别成功率获得。

**[0053]** 在本发明实施例中，NVM 210 是新一代的非易失性内存。NVM 210 的存取速度与传统的易失性内存（例如，动态随机存取存储器 DRAM 或静态随机存取存储器 SRAM）的存取速度相当。此外，NVM 210 具有半导体产品的可靠性，使用寿命较长，能够实现按字节（Byte）寻址，将数据以位（bit）为单位写入存储介质中。因此，NVM 210 能够被挂在内存总线上，作为内存被 CPU 10 直接访问。需要说明的是，NVM 210 与传统的易失性内存不同，是非易失（Non-Volatile）的，当终端设备 100 关闭电源后，NVM 210 中的信息依然存在。在本发明实施例中，NVM 210 可以包括相变存储器（Phase Change Memory，PCM）、阻变存储器(Resistive Random Access Memory, RRAM)、磁性随机存储器（Magnetic Random Access Memory，MRAM）和铁电式随机存储器（Ferroelectric Random Access Memory，FRAM）等为代表的下一代非易失性存储器（Non-Volatile Memory, NVM）。具体的，由于自旋转移矩磁随机存取存储器(spin-transfer-torque magnetic RAM，STT-MRAM)的使用寿命较长且功耗较低，并且，由于 STT-MRAM 的写入成功率与写入电流关系较大，在本发明实施例中，NVM 210 可以为 STT-MRAM。

**[0054]** NVM 210 可以包括 NVM 控制器 2102、第一存储区域 2104 和第二存储区域 2106。NVM 控制器 2102 用于访问第一存储区域 2104 和第二存储区域 2106。例如，NVM 控制器 2102 可以将数据写入第一存储区域 2104 和第二存储区域 2106，或者从第一存储区域 2104 和第二存储区域 2106 中读取数据。实际应用中，在 NVM 控制器 2102 中，可以包括处理器、特定集成电路（Application Specific Integrated Circuit，ASIC）或者被配置成实施本发明实施例的一个或多个集成电路。在 NVM 控制器 2102 中还可能包括缓存、通信接口等，在此不对 NVM 控制器 2012 的具体结构进行限定。

**[0055]** 第一存储区域 2104 和第二存储区域 2106 可以是由多个存储单元构成的存储区域。在本发明实施例中，存储单元是指用于存储数据的最小存储介质单元，每个存储单元用于存储 1 比特（bit）的数据。例如，存储单元可以包括相变存储单元、磁性存储单元、阻变存储单元等非易失性存储单元。在本发明实施例中，以 NVM 210 为 STT-MRAM 为例，第一

存储区域2104和第二存储区域2106可以是由多个磁性存储单元构成的存储阵列。本领域技术人员可以知道，每个磁性存储单元包括两个磁性层和一个隧道层。其中，一个磁性层的电磁方向固定，另一磁性层的电磁方向可以通过外部的电磁场来改变。当两个磁性层的方向一样时，该磁性存储单元的电阻低，用于代表数据“0”；当两个磁性层的方向相反时，该磁性存储单元的电阻为高，用于代表数据“1”。通常，本领域技术人员将能够通过外部电磁场改变电磁方向的磁性层称为自由层。在本发明实施例中，可以通过将自旋偏振电流通过磁性存储单元来改变自由层的磁场方向。需要说明的是，在本发明实施例中，第一存储区域 2104和第二存储区域2106并不一定是连续的地址空间。并且，在NVM 210中除了第一存储区域2104和第二存储区域2106外，还可以包括用于存储其他数据的存储空间（图中未示出），在此不做限定。

**[0056]** 本领域技术人员可以知道，与传统内存相比，非易失性内存基本不存在静态功耗，然而对非易失性内存执行读操作和写操作所造成的能量开销(也可称为动态功耗)较大。其中，静态功耗是指未对非易失性内存执行读操作和写操作期间所造成的能量开销。为了达到降低终端设备的系统功耗的目的，可以通过降低NVM的动态功耗来实现。具体的，可以通过控制写操作过程中的写入电流的大小来控制NVM的动态功耗。然而，本领域技术人员可以知道，在向磁性存储单元写数据的过程中，写入电流的强度要超过阈值电流才能保证磁性存储单元的电阻状态的翻转，因此，写入成功率与写入电流的大小也密切相关。在实现本发明的过程中，发明人发现，对于部分图像、视频等应用，数据中的低位部分在存储过程中出现的错误对识别成功率的影响较小。为了在降低写入功耗的同时不影响图像数据识别的成功率，在本发明实施例中，NVM 210 采用了通过不同的写入电流结合存储的方式来存储图像数据。根据这种方式，NVM控制器2102可以通过控制写入电流将经过降维处理模块205降维处理后的第一图像数据中各个数值的低位部分和高位部分分别写入第一存储区域 2104和第二存储区域2106。具体的，在本发明实施例中，第一存储区域2104的写入电流I低于第二存储区域2106的写入电流$I_s$。例如，第一存储区域2104的写入电流可以为第一电流I，第二存储区域2106的写入电流$I_s$可以为2I。本领域技术人员可以知道，NVM控制器2102可以通过控制写入电压来控制写入电流的大小。

**[0057]** 图6为本发明实施例提供的一种NVM 210的硬件结构示意图。如图6所示，第一存储区域2104和第二存储区域2106包括多个磁性存储单元610构成的存储阵列。NVM控制器2102可以通过控制第一电压V来控制第一电流I，NVM控制器2102可以通过控制第二电压 Vs 来控制第二电流 $I_s$。同一列的磁性存储单元 610 可以连接一个多路选择器

(multiplexer, MUX) 605。NVM 控制器 2102 可以通过控制信号控制多路选择器 605 输出第一电压 V 还是输出第二电压 Vs，以实现选择通过第一电流 I 将降维后的第一图像数据中的各数值的低 ω 位写入第一存储区域 2104 或通过第二电流 $I_s$ 将各数值的高(N-ω)位写入第二存储区域 2106 的目的。其中，N 为每一个数值所占的比特位，ω 为设置的宽度参数。例如，若待识别的图像数据的数值为 64 bit，则可以按照第一电流 I 将该数值的低 16bit 写入第一存储区域 2104，按照第二电流 $I_s$ 将该数值的高 48bit 写入第二存储区域 2106。为了描述方便，在本发明实施例中，将 ω 称为宽度参数。实际应用中，ω 的值以及第一电流 I 的值均需要根据终端设备 100 的系统功耗以及设置的图像识别成功率来确定。可以理解的是，待识别的图像数据的类型不同，对图像识别成功率的要求也不相同，设置的宽度参数 ω 以及第一电流 I 的值也就不同，其中，ω 的值为正整数。

**[0058]** 可以理解的是，图 6 仅仅是为了阐述图像识别加速器 20 中的 NVM 210 如何将图像数据进行分区存储而对 NVM 210 中的部分结构做出的示意性图示。实际应用中，多路选择器 MUX 605 可能并不直接连接磁性存储单元 610，而是通过 STT-MRAM 中的写装置（图 6 中未示出）将数据写入磁性存储单元 610。此外，实际应用中，也可以为多列磁性存储单元 610 设置一个 MUX 605，或者为一行或多行磁性存储单元 610 设置一个 MUX 605。在此不对 MUX 605 的数量以及 MUX 605 与磁性存储单元 610 的连接关系进行限制，只要能够实现将图像数据中的数值的不同部分按照不同的电流分别写入不同的磁性存储单元 610 即可。

**[0059]** 图像匹配模块 215，用于确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据，并输出匹配结果。具体的，结合图 4，在步骤 410 中，图像匹配模块 215 可以确定所述 NVM 210 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据，以获得所述降维后的第一图像数据与 NVM 210 中存储的图像库中的图像数据的匹配结果。例如，图像匹配模块 215 可以分别从第一存储区域 2104 和第二存储区域 2106 读取降维后的第一图像数据，并将降维后的第一图像数据直接与 NVM 210 中存储的图像库中的图像数据进行匹配，以判断是否能够成功识别该第一图像数据。可以理解的是，为了识别图像，在 NVM 210 中需要预先存储包含有至少一个图像数据的图像库。在本发明实施例中，图像匹配模块 215 可以是逻辑电路或 ASIC 芯片。例如，图像匹配模块 215 可以通过逻辑电路或 ASIC 芯片将降维后的第一图像数据与图像库中的图像数据按照匹配追踪(Matching Pursuits，MP)算法进行计算，从而确定所述 NVM 中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据，以获得匹配结果。可以理解的是，图像库中的图像数据也可以是经过与第一图像数据相同的处理方法存储于 NVM 210 中的图

像数据。

**[0060]** 需要说明的是，本发明实施例并不对图像匹配模块 215 的具体实现形式进行限制，只要能够实现图像数据的匹配过程即可。并且，本发明实施例也不对具体的匹配算法进行限定，实际应用中，可以采用正交匹配追踪(Orthogonal Matching Pursuit，OMP)算法，也可以采用其他匹配算法，在此，不对图像匹配模块 215 采用的匹配算法进行限定。实际应用中，匹配模块 215 获得匹配结果后，可以将匹配结果返回给 CPU 或者将匹配结果发生给其他数据处理模块，在此不进行限定。

**[0061]** 在本发明实施例中，由于 NVM 210 将降维后的第一图像数据中的各个数值的不同部分按照不同的电流分别写入第一存储区域 2104 和第二存储区域 2106，且第一电流 I 小于第二电流 Is，因此，按照第一电流 I 将数据存储于第一存储区域 2104 与按照第二电流 Is 将数据存储于第二存储区域 2106 相比更节省系统功耗。本领域人员可以知道，通常，写入电流越低，存储的数据出现错误的机会将越大。或者换一种表达方式，随着写入电流的降低，图像数据的识别成功率将会降低。因此，现有技术中通常会先将存储的图像数据通过解凸优化等恢复方式进行恢复后再进行图像识别。在本发明实施例中，由于设置的宽度参数 ω 以及第一电流 I 是根据终端设备 100 的系统功耗和设置的第一图像识别成功率获得的，从而使得存储于第一存储区域 2104 的数值中的低位部分在存储过程中出现的错误对识别成功率的影响较小。因此，图像匹配模块 215 在实现图像数据匹配的过程中，并不需要将图像数据恢复后进行匹配，而可以直接将存储于 NVM 210 中的降维后的第一图像数据与图像库中的图像数据进行匹配。通过本发明实施例提供的这种图像识别方式，能够在节省系统功耗的情况下满足设置的图像识别成功率，保证存储的图像数据的准确性。

**[0062]** 为了使终端设备 100 能够满足各种类型的图像数据的识别需求，并能够在节省系统功耗的情况下满足设置的图像识别成功率，在本发明实施例提供的图像识别加速器 20 中还可以设置有统计模块 225 以及参数调整模块 220。如图 7 所示，图 7 为本发明实施例提供的又一种图像识别加速器 20 的结构示意图。如图 7 所示，在图 3 所示的结构基础上，参数调整模块 220 分别连接降维处理模块 205 以及 NVM 210。统计模块 225 分别与匹配模块 215 以及参数调整模块 220 连接。下面将结合图 8 所示的又一种图像识别方法对图 7 所示的图像识别加速器 20 中的各个器件的结构和工作原理进行详细的介绍。

**[0063]** 统计模块 225 用于统计在预设的统计期间内图像匹配模块 215 输出的匹配结果，以获得统计的图像识别成功率。从而参数调整模块 220 能够根据统计模块 225 统计的图像识

别成功率以及设置的第二图像识别成功率判断是否需要调整图像识别参数。具体的，如图 8 所示，在步骤 800 中，统计模块 225 可以统计在预设的统计期间内图像匹配模块 215 输出的匹配结果，获取所述统计的图像识别成功率。可以理解的是，统计模块 225 获得的图像识别成功率是根据多个图像数据的识别结果获得的。实际应用中，统计模块 225 可以是计数器等器件，在此不对统计模块 225 的具体实现形式进行限制。

**[0064]** 可以理解的是，图 7 仅仅是对统计模块 225 的一种结构的示意，实际应用中，还可以将统计模块 225 单独设置于终端设备 100 中，也可以将统计模块 225 设置在 CPU 10 中，或者将统计模块 225 设置在与匹配模块 215 连接的其他设备中，本发明实施例不对统计模块 225 设置的具体位置进行限定。

**[0065]** 参数调整模块 220，用于如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数 γ、宽度参数 ω 以及第一电流 I。为了描述方便，在本发明实施例中，可以将降维参数 γ、宽度参数 ω 以及第一电流 I 统称为图像识别参数。具体的，参数调整模块 220 可以根据统计模块 225 统计的图像识别成功率与设置的第二图像识别成功率的差值的绝对值来判断是否需要调整图像识别参数的取值。其中，第二图像识别成功率为重新设置的图像识别成功率，第二图像识别成功率与前述的第一图像识别成功率不同。可以理解的是，第二图像识别成功率可以预先从 CPU 10 获得。结合图 8 所示，若在步骤 805 中，参数调整模块 220 确定统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则在步骤 810 中，参数调整模块 220 可以根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数 γ、宽度参数 ω 以及第一电流 I。

**[0066]** 为了清楚阐述参数调整模块 220 具体如何调整图像识别参数，下面将结合图 9 所示的参数调整方法流程图对参数调整模块 220 如何平衡系统功耗和图像识别成功率，以获得合适的图像识别参数进行描述。图 9 为本发明实施例提供的一种参数调整方法流程图。在本发明实施例中，以需要将图像识别成功率从第一图像识别成功率调整为第二图像识别成功率为例进行描述。如图 9 所示，该参数调整方法可以包括下述步骤。

**[0067]** 在步骤 900 中，参数调整模块 220 分别逐渐调整降维参数 γ、所述宽度参数 ω 或第一电流 I 的取值以降低所述终端设备的系统功耗 E，并分别通过统计模块 225 获得调整后的图像识别成功率。根据前面的描述可以知道，第一电流 I 的值越小，终端设备 100 存储图像数据时的动态功耗越小，则终端设备 100 的系统功耗越小。随着宽度参数 ω 的值增大，按

照第一电流 I 存储于第一存储区域 2104 的数据越多，则终端设备 100 的系统功耗越小。降维参数 γ 的值越小，降维后的图像数据的数据量越少，终端设备 100 的系统功耗也会越小。终端设备的系统功耗 E 的值与 $\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比，其中，$I_s$ 为设置的标准写入电流，或者也可以被称为是安全电流，按照 $I_s$ 将数据写入 NVM 210 中时，可以保证图像数据的准确度。通常，$I_s$ 的值可以根据 NVM 210 的物理参数获得，在使用 NVM 210 的过程中，NVM 210 的 $I_s$ 不会改变。可以理解的是，不同工艺制造的 NVM 210 因其物理参数不同，因此 $I_s$ 也可能不同。所述第二图像数据是用数值表示的第二图像的各像素的灰度值的集合，所述第二图像数据可以包括多个数值。实际应用中，在调整图像识别参数的过程中，参数调整模块 220 可以分别通过逐渐增加宽度参数 ω 的值、或逐渐降低降维参数 γ 或逐渐第一电流 I 的值的方式降低终端设备 100 的系统功耗 E，并记录下调整的过程中的各个参数的值、系统功耗以及获得的图像识别成功率。记录的形式可以是图 10(a)所示的表格形式，也可以是如图 10(b)所示的图形的形式，还可以是其他记录形式。其中，在图 10（b）中仅对如何采用图表的形式记录 ω 和 I 的调整对图像识别成功率的值的改变进行了示例。可以看出，用图形的形式比用表格记录更加直观。在本发明实施例中，图像识别成功率也可以被称为服务质量 QoS。可以理解的是，图像识别成功率可以采用确定的图像识别参数识别多个图像数据的试验来获得。系统功耗可以根据 $\gamma((N-\omega)* I_s^2 +\omega*I)$的公式计算获得，可以理解的是，系统功耗的值可以是一个估计值。

**[0068]** 具体的，在调整参数的过程中，当调整了一次降维参数 γ、所述宽度参数 ω 或第一电流 I 的值后，可以通过图 3 所示的图像识别加速器识别多个实验数据，以获得根据调整的降维参数 γ、所述宽度参数 ω 或第一电流 I 的值识别多个实验数据的识别成功率，并根据 $\gamma((N-\omega)* I_s^2 +\omega*I)$计算获得每次调整参数后的系统功耗，以得到图 10（a）所示的多组参数值以及对应的系统功耗和图像识别成功率的值。可以理解的是，在本发明实施例中，实验数据也为图像数据。例如，以调整过程中调整的参数值为图 10（a）中的 γ3、ω3 和 I3 为例。参数调整模块 220 在将调整降维参数、所述宽度参数或第一电流的值调整为第一组参数值: γ3、ω3 和 I3 后，参数调整模块 220 可以根据 $\gamma((N-\omega)* I_s^2 +\omega*I)$的公式计算获得与所述第一组参数值对应的系统功耗 E5。并且，参数调整模块 220 可以将调整后的参数值 γ3 发送给降维处理模块 205，将 ω3 和 I3 发送给 NVM 210。降维处理模块 205、NVM 210 以及图像识别模块分别根据调整后的参数值 γ3、ω3 和 I3 按照图 4 所示的方法对实验数据进行识别，以得到相应的系统功耗和图像识别成功率。具体的，降维处理模块 205 根据接收的降维参数值 γ3 对实验数据进行降维处理。NVM 210 根据 I3 将降维后的实验数据中的低 ω3 位存储在第一存

储区域2104中，并根据Is将降维后的实验数据中的高（N-ω3）位存储于第二存储区域2106中。图像匹配模块215可以分别从第一存储区域2104和第二存储区域2106读取降维后的实验数据，并将降维后的实验数据直接与NVM 210中存储的图像库中的图像数据进行匹配，以判断是否能够成功识别该实验数据。按照这种方式，按照γ3、ω3和I3对多个实验数据进行识别后，可以获得该组参数值对应的图像识别成功率Qos5。如果Qos5不满足设定的第二图像识别成功率的要求，可以继续调整降维参数γ、所述宽度参数ω或第一电流I的值，再根据调整后的参数值对实验数据按照图4所示的方法进行识别。从而在调整参数过程中，能够按照这种方式获得每次调整参数值后的图像识别成功率以及系统功耗。例如，根据这种方式可以得到如图10（a）所示的多组参数值以及对应的系统功耗和图像识别成功率。

**[0069]** 实际应用中，由于宽度参数ω的值为整数，因此在调整过程中，为了方便调整，可以优先调整宽度参数ω的值，并以调整后的宽度参数ω的值为依据分别调整γ和I的值，以使调整后的参数值对实验数据识别后能够满足设定的图像识别成功率（例如第二图像识别成功率）的要求。本发明实施例不对参数值的具体调整顺序进行限定。可以理解的是，在调整参数的过程中，当调整参数值后，可以通过调整后的参数值对预设数量的实验数据进行识别以获得图像识别成功率。在本发明实施例中，可以将参数调整过程中识别多个实验数据的识别成功率称为调整后的识别成功率。可以理解的是，在本发明实施例中，可以预先设置图像数据的实验库，在实验库中存储有实验用的图像数据，以用于在调整参数过程中作为实验数据使用。需要说明的是，在本发明实施例中，除了图10（a）所示的列表中的表头部分（图10（a）中第一行）的ω、γ、I、E及Qos是用于表示参数外，表中除第一行之外的其他部分中的ω、γ、I、γ1、ω1、I2、E1、Qos1等均用于表示具体的参数值，本发明实施例中其他部分的ω、γ、I、γ'、ω'以及I'均用于表示具体的参数值。换一种表达方式，在本发明实施例中，如无特别说明，ω和ω'均用于表示宽度参数的值，γ和γ'均用于表示降维参数的值，I和I'均用于表示第一电流的值。

**[0070]** 在步骤905中，参数调整模块905确定在调整后的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值不大于所述预设阈值时所述终端设备的最小功耗E'。可以理解的是，在步骤900所示的调整参数的过程中，可以获得与调整的参数对应的多个图像识别成功率以及多个系统功耗。本领域技术人员可以理解的是，降维参数γ的值越小，降维后的图像数据的数据量越小，出错机会更小，但降维后的图像数据中每个数值包含的信息量更大。因此，实际应用中，会存在降维参数γ减小而图像识别成功率更高的情况。从而，在选择参数时，需要考虑降维参数γ与图像识别成功率的折衷。

**[0071]** 在本发明实施例中，可以将与设置的第二图像识别成功率的差值的绝对值不大于预设阈值的图像识别成功率都作为是满足第二图像识别成功率的要求的图像识别成功率。例如，若第二图像识别成功率为 90%，预设阈值为 2%，则在 88%-92%之间的图像识别成功率均可以认为是满足第二图像识别成功率要求的图像识别成功率。在本步骤中，可以在记录的多个图像识别成功率中确定满足第二图像识别成功率要求的至少一个图像识别成功率。并在确定的至少一个图像识别成功率对应的多个系统功耗中确定最小的系统功耗 E'。

**[0072]** 在步骤 910 中，参数调整模块 220 选择在满足所述最小系统功耗 E'时获得最大图像识别成功率的降维参数 γ、所述宽度参数 ω 以及第一电流 I 分别作为所述调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。可以理解的是，在步骤 905 中确定的最小系统功耗 E'对应的满足第二图像识别成功率要求的图像识别成功率可以有多个。因此，在步骤 910 中，参数调整模块 220 可以选择满足所述最小系统功耗 E'时获得最大图像识别成功率的降维参数 γ、所述宽度参数 ω 以及第一电流 I 作为调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。例如，在第一种情形下，宽度参数 ω 增加 1bit，获得的图像识别成功率为 88%，系统功耗 E'为 10w。在第二种情形下，降维参数 γ 减少 0.5，获得的图像识别成功率为 90%，系统功耗 E'也为 10w。在第三种情形下，电流 I 减少 500μA，获得的图像识别成功率为 92%，系统功耗 E'也为 10w。则可以将第三种情形下的降维参数 γ、所述宽度参数 ω 以及第一电流 I 作为调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。

**[0073]** 可以理解的是，本发明实施例仅仅对参数调整模块 220 调整图像识别参数的过程进行了一个简单的示例，实际应用中，还可以组合调整上述三个参数，例如，可以同时将宽度参数 ω 增加 1bit 并将降维参数 γ 减少 0.5。在本发明实施例中不对具体的调整形式进行限定，只要调整上述三个图像识别参数中的至少一个参数即可。实际应用中，参数调整模块 220 可以根据贪心算法来确定调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。

**[0074]** 需要说明的是，实际应用中，统计模块 225 还可以位于终端设备 100 的 CPU 10 中，则在这种情形下，参数调整模块 220 可以根据 CPU10 的指示调整图像识别参数。如图 11 所示，图 11 为本发明实施例提供的又一种终端设备的结构示意图。在图 11 所示的结构中，统计模块 225（图 11 中未示出）可以位于 CPU 10 中， CPU 10 可以根据图像匹配模块 215 在预设的统计期间发送的匹配结果统计图像识别加速器 20 的图像识别成功率。如果统计的图像识别成功率与所述第二图像识别成功率之间的差值大于预设阈值，则 CPU 10 可以向参数调整模块 220 发送参数调整指令，以指示参数调整模块 220 调整图像识别参数。所述参数调整指令中包含有所述第二图像识别成功率。换一种表达方式，在图 11 所示的结构中，由

CPU10 和参数调整模块 220 共同完成调整图像识别参数的功能，具体的，可以由 CPU 10 执行图 8 中所示的步骤 800-805 的动作，并指示参数调整模块 220 执行步骤 810 的动作。

**[0075]** 在又一种情形下，还可以由终端设备 100 中的 CPU 10 完成调整图像识别参数的功能。如图 12 所示，图 12 为本发明实施例提供的又一种终端设备结构示意图。图 12 在图 11 的基础上减少了参数调整模块 220，图 11 中的参数调整模块 220 的功能改由 CPU 10 执行。具体的，在图 12 所示的终端设备 100 的结构示意图中，图像匹配模块 215 可以将匹配结果反馈给 CPU 10，CPU 10 可以根据匹配结果统计预设统计期间图像识别加速器 20 识别图像数据的图像识别成功率。CPU 10 可以根据统计的图像识别成功率以及重新设置的第二图像识别成功率确定是否需要调整图像识别参数。当 CPU 10 确定统计的图像识别成功率与设置的第二图像识别成功率的差值的绝对值大于预设阈值时，CPU 10 可以根据第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个图像识别参数：降维参数 γ、宽度参数 ω 以及第一电流 I，并将调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'分别发送给降维处理模块 205 和 NVM 210。换一种表达方式，在图 12 所示的结构中，CPU 10 可以执行图 8 中所示的步骤 800-810 的方法。各器件具体功能的描述可以参见前述实施例的相关描述，在此不再赘述。可以理解的是，在图 12 所示的结构中，当 CPU 10 调整图像识别参数时，可以采用指令的形式向图像识别加速器 20 中的降维处理模块 205 和 NVM 210 发送调整的降维参数 γ'、宽度参数 ω'以及第一电流 I'，从而控制降维处理模块 205 和 NVM 210 按照调整后的图像识别参数对图像数据进行识别。

**[0076]** 本领域技术人员可以理解的是，虽然上述实施例是以将图像识别成功率从第一图像识别成功率调整为第二图像识别成功率为例对本发明实施例提供的图像识别加速器实现参数调整的过程进行描述，但上述示例只是为了对图像识别加速器能够实现的参数调整功能，从而获得满足条件的图像识别参数（包括降维参数、宽度参数以及第一电流）值的一个示例。可以理解的是，在本发明实施例中，用于对第一图像数据进行识别的降维参数 γ、宽度参数 ω 以及第一电流 I 也是依据上述参数调整方式调整后获得的。换一种表达方式，上述参数调整方法是对如何通过调整参数值以获得满足系统功耗以及图像识别成功率的需求的图像识别参数值的一种方法的描述，实际应用中，当需要调整参数时，均可以按照上述参数调整方法获得满足需求的参数值。

**[0077]** 在获得调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'后，图像识别加速器 20 可以根据调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'对后续需要识别的第二图像数据进行图像识别。具体的，如图 8 所示，在步骤 815 中，降维处理模块 205 可以根据调整后

的降维参数γ'降低第二图像数据的维度。在步骤820中，NVM 210可以将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于NVM 210的第二存储区域，其中，所述 I'小于所述 $I_s$。在步骤825中，图像匹配模块215可以确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。具体的，图像匹配模块215可以将降维后的第二图像数据与NVM 210中存储的图像库中的图像数据进行比较，以获得所述降维后的第二图像数据与NVM 210中存储的图像库中的图像数据的匹配结果。可以理解的是，图像识别加速器根据调整后的降维参数γ'、宽度参数ω'以及第一电流I'对第二图像数据进行图像识别的过程与前述的根据降维参数γ、宽度参数ω以及第一电流I对第一图像数据进行图像识别的过程类似，具体描述可以参见前面的描述，在此不再赘述。

**[0078]** 可以理解的是，本发明实施例中的参数调整方法只在需要调整参数时才会触发调整参数。换一种表达方式，当需要调整图像识别参数的值时，图像识别加速器可以触发停止接收待识别的图像数据（也可称为业务数据），而是按照图 9 所示的方式通过调整参数值并识别实验数据的方式获得满足需求的图像识别参数的值。在获得满足图像识别需求的参数值后，再将满足需求的参数值分别发送给降维处理模块205以及NVM 210，使得降维处理模块205 、NVM 210以及图像匹配模块215能够根据调整获得的图像识别参数值对待识别的图像数据（例如第一图像数据和第二图像数据）进行图像识别。

**[0079]** 为了更清晰的描述本发明实施例提供的图像识别加速器20如何识别图像数据，下面将结合图13所示的图像识别方法的信令图对图7所示的图像识别加速器20的工作过程进行简单的描述。在本发明实施例中，仍然以识别第一图像数据为例进行描述。如图 13 所示。降维处理模块205接收CPU 10发送的待识别的第一图像数据1300后，降维处理模块205可以根据参数调整模块220设置的降维参数γ降低第一图像数据1300的维度。其中降维处理模块205可以采用伯努利矩阵对第一图像数据1300进行降维，以便能够基于稀疏表示的随机映射的方式对第一图像数据1300进行降维。NVM 210接收到降维后的第一图像数据1305后，NVM 210可以根据参数调整模块220设置的宽带参数ω以及第一电流I将降维后的第一图像数据的每一个数值中的低ω位按照第一电流I存储于第一存储区域2104，并将降维后的第一图像数据的每一个数值中的高N-ω位按照第二电流Is存储于第二存储区域2106。图像匹配模块215可以基于NVM210存储的图像库中的图像数据对降维后的第一图像数据1305进行识别，确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据

相匹配的图像数据，并输出匹配结果。为了图示方便，在图 13 中，将图像库中的图像数据以及降维后的第一图像数据 1305 统称为待比较的图像数据 1310。一方面，在图 13 中，图像匹配模块 215 可以向 CPU 10 输出第一图像数据 1305 的识别结果。另一方面，统计模块 225 可以统计图像匹配模块 215 的图像匹配结果，从而获得统计期间内的图像识别成功率 1320，参数调整模块 220 从而可以根据统计模块 225 获得的图像识别成功率 1320 与设置的第二图像识别成功率确定是否需要调整图像识别参数。当参数调整模块 220 确定需要调整图像识别参数时，参数调整模块 220 可以根据图 9 所示的方法对图像识别参数进行调整，并向降维模块 205 和 NVM 210 分别输出调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'。从而，降维处理模块 205、NVM 210 以及图像匹配模块 215 能够根据调整后的降维参数 γ'、宽度参数 ω'以及第一电流 I'对后续的第二图像数据进行识别。

**[0080]** 可以理解的是，图 13 仅仅是对本发明实施例提供的一种终端设备 100 的信令示意图，对其他实施例提供的图像识别加速器 20 或终端设备 100 的工作过程可以参见图 13 以及前述实施例的描述。在此不再赘述。

**[0081]** 本发明实施例提供的终端设备，通过图像识别加速器识别图像，减少了 CPU 数据处理量，也减少了 CPU 与内存的数据交互，能够减少 CPU 的负担，且能够减少内存带宽对图像数据识别应用的限制，提高图像数据的识别速度。并且，在本发明实施例提供的终端设备中，图像识别加速器可以根据基于稀疏表示的随机映射的方式降低待识别的图像数据的维度，然后将降维后的图像数据按照不同的电流写入图像识别加速器的 NVM 中的不同存储区域。由于设置的降维参数 γ、宽度参数 ω 以及第一电流 I 均是根据所述终端设备的系统功耗和设置的图像识别成功率获得的，因此能够在降低终端设备的系统功耗的基础上保证图像识别的准确性。

**[0082]** 本发明实施例还提供一种数据处理的计算机程序产品，包括存储了程序代码的计算机可读存储介质，所述程序代码包括的指令用于执行前述任意一个方法实施例所述的方法流程。本领域普通技术人员可以理解，前述的存储介质包括：U 盘、移动硬盘、磁碟、光盘、随机存储器 (Random-Access Memory，RAM)、固态硬盘（Solid State Disk，SSD）或者非易失性存储器（non-volatile memory）等各种可以存储程序代码的非短暂性的（non-transitory）机器可读介质。

**[0083]** 需要说明的是，本申请所提供的实施例仅仅是示意性的。所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中没有详述的部分，可以参见其他实施例的相关描述。在本发明实施例、权利要求以及附图中揭示的特征可以独立存在也可以组合存在。在本发明实施例中以硬件形式描述的特征可以通过软件来执行，反之亦然。在此不做限定。

# 权 利 要 求

1、一种应用于终端设备中用于识别图像的图像识别加速器，其特征在于，包括:

降维处理模块，用于接收降维参数γ，根据接收的降维参数γ降低第一图像数据的维度，其中，降维后的第一图像数据包括多个数值，所述降维参数γ是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的;

非易失性内存NVM，用于接收宽度参数ω和第一电流I，将降维后的第一图像数据的每一个数值的低ω位按照设置的第一电流I存储于所述NVM的第一存储区域，将降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，所述第一电流I小于所述第二电流$I_s$，所述宽度参数ω以及所述第一电流I是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的;

图像匹配模块，用于确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

2、根据权利要求1所述的图像识别加速器，其特征在于，还包括:

参数调整模块，用于根据设置的所述第一图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数、宽度参数以及第一电流的值，以获得调整后的降维参数γ、宽度参数ω以及第一电流I，并将调整后的所述降维参数γ发送给所述降维处理模块，将调整后的所述宽度参数ω以及调整后的所述第一电流I发送给所述NVM。

3、根据权利要求2所述的图像识别加速器，其特征在于，所述参数调整模块具体用于:

分别调整所述降维参数、所述宽度参数或所述第一电流的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，每个调整后的图像识别成功率对应于每个调整后的系统功耗;

确定调整后的每个图像识别成功率与所述第一图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗;

选择在满足所述最小系统功耗时获得最大图像识别成功率的降维参数、宽度参数以及第一电流的值分别作为调整后的所述降维参数γ、所述宽度参数ω以及所述第一电流I，

并将调整后的所述降维参数γ发送给所述降维处理模块，将调整后的所述宽度参数ω以及所述第一电流I发送给所述NVM。

4、根据权利要求1所述的图像识别加速器，其特征在于：

所述参数调整模块，还用于如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，以获得调整后的降维参数γ'、宽度参数ω'以及第一电流I'，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

所述降维处理模块，还用于根据调整后的所述降维参数γ'降低第二图像数据的维度；

所述非易失性内存NVM，还用于将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述ω'为调整后的宽度参数，所述 I'小于所述$I_s$；

所述图像匹配模块，还用于确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

5、根据权利要求1-4任意一项所述的图像识别加速器，其特征在于，还包括：

统计模块，用于统计在预设的统计期间内所述图像匹配模块输出的匹配结果，获取所述统计的图像识别成功率。

6、根据权利要求1-5任意一项所述的图像识别加速器，其特征在于，所述降维处理模块具体用于：

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的值根据设置的降维参数γ确定，γ=n/m。

7、根据权利要求2或3所述的图像识别加速器，其特征在于，所述参数调整模块具体用于：

如果统计的图像识别成功率与所述第二图像识别成功率之间的差值大于预设阈值，则分别调整降维参数γ、所述宽度参数ω或第一电流I的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比，每个调整后的图像识别成功率对应于每个调整后的系统功耗；

确定调整后的每个图像识别成功率与所述第二图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值时的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗E'；

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数、宽度参数以及第一电流的值分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'，并将所述调整后的降维参数γ'发送给所述降维处理模块，将所述调整后的宽度参数ω'以及第一电流I'发送给所述NVM。

8、根据权利要求6所述的图像识别加速器，其特征在于，所述二进制类型的矩阵包括伯努利映射矩阵。

9、一种终端设备，其特征在于，包括CPU和图像识别加速器，其中：

所述CPU，用于向所述图像识别加速器发送待识别的第一图像数据；

所述图像识别加速器，用于根据降维参数γ降低所述第一图像数据的维度，其中，降维后的第一图像数据包括多个数值，所述降维参数γ是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的；

将所述降维后的第一图像数据的每一个数值的低ω位按照第一电流I存储于NVM的第一存储区域，将所述降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，ω为宽度参数，所述I小于所述$I_s$，所述宽度参数ω以及所述第一电流I是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

10、根据权利要求9所述的终端设备，其特征在于，所述图像识别加速器还用于：

根据设置的所述第一图像识别成功率以及所述终端设备的系统功耗调整下述至少

一个参数：降维参数、宽度参数以及第一电流的值，以获得所述降维参数γ、所述宽度参数ω以及所述第一电流I。

11、根据权利要求10所述的终端设备，其特征在于，所述图像识别加速器具体用于：

分别调整所述降维参数、所述宽度参数或所述第一电流的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，每个调整后的图像识别成功率对应于每个调整后的系统功耗；

确定调整后的每个图像识别成功率与所述第一图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗；

选择在满足所述最小系统功耗时获得最大图像识别成功率的降维参数、所述宽度参数以及第一电流的值分别作为所述降维参数γ、所述宽度参数ω以及所述第一电流I。

12、根据权利要求9所述的终端设备，其特征在于：

所述图像识别加速器，还用于如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，以获得调整后的降维参数γ'、宽度参数ω'以及第一电流I'，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

所述CPU，还用于向所述图像识别加速器发送第二图像数据；

所述图像识别加速器，还用于：

根据调整后的所述降维参数γ'降低第二图像数据的维度；

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的所述第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_S$存储于所述NVM的第二存储区域，其中，ω'为调整后的宽度参数，所述I'小于所述$I_S$；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

13、根据权利要求9所述的终端设备，其特征在于：

所述CPU，还用于统计在预设的统计期间内所述图像识别加速器输出的匹配结果，获取统计的图像识别成功率；

确定所述统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值；

所述图像识别加速器，还用于根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I的值，以获得调整后的降维参数γ'、宽度参数ω'以及第一电流I'，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

所述CPU，还用于向所述图像识别加速器发送第二图像数据；

所述图像识别加速器，还用于根据调整后的所述降维参数γ'降低第二图像数据的维度；

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的所述第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述ω'为调整后的宽度参数，所述 I' 小于所述$I_s$；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

14、根据权利要求9所述的终端设备，其特征在于，所述CPU还用于：

统计在预设的统计期间内所述图像识别加速器输出的匹配结果，获取所述统计的图像识别成功率；

如果统计的图像识别成功率与设置的第二图像识别成功率之间的差值的绝对值大于预设阈值，则根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，以获得调整后的降维参数γ'、宽度参数ω'以及第一电流I'，其中，所述第二图像识别成功率与所述第一图像识别成功率不同；

向所述图像识别加速器发送第二图像数据；

所述图像识别加速器，还用于：

根据调整后的所述降维参数γ'降低第二图像数据的维度；

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的所述第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述

第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述ω'为调整后的宽度参数，所述 I' 小于所述$I_s$;

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

15、根据权利要求9-14任意一项所述的终端设备，其特征在于，所述图像识别加速器具体用于:

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的值根据设置的降维参数γ确定，γ=n/m。

16、根据权利要求10或13所述的终端设备，其特征在于，所述图像识别加速器具体用于:

分别调整降维参数γ、所述宽度参数ω或第一电流I的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比，每个调整后的图像识别成功率对应于每个调整后的系统功耗;

确定调整后的每个图像识别成功率与所述第二图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值时的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗E';

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数、宽度参数以及第一电流I的值分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

17、根据权利要求14所述的终端设备，其特征在于，所述CPU具体用于:

分别调整降维参数γ、所述宽度参数ω或第一电流I的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比，每个调整后的图像识别成功率对应于每个调整后的系统功耗;

确定调整后的图像识别成功率与所述第二图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值时的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗E';

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数、宽度参数以及第一电流分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

18、根据权利要求15所述的终端设备，其特征在于：所述二进制类型的矩阵包括伯努利映射矩阵。

19、一种应用于终端设备的图像识别方法，其特征在于，所述方法由所述终端设备中的图像识别加速器执行，所述方法包括：

根据降维参数γ降低第一图像数据的维度，其中，降维后的第一图像数据包括多个数值，所述降维参数γ是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的；

将所述降维后的第一图像数据的每一个数值的低ω位按照设置的第一电流I存储于所述图像识别加速器中的NVM的第一存储区域，将所述降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域，其中，N为每一个数值所占的比特位，ω为宽度参数，所述I小于所述$I_s$，所述宽度参数ω以及第一电流I是根据所述终端设备的系统功耗和设置的第一图像识别成功率获得的；

确定所述NVM中存储的图像库中是否包含有与所述降维后的第一图像数据相匹配的图像数据。

20、根据权利要求19所述的图像识别方法，其特征在于，还包括：

根据设置的所述第一图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数、宽度参数以及第一电流的值，以获得调整后的降维参数γ、宽度参数ω以及第一电流I。

21、根据权利要求20所述的图像识别方法，其特征在于，所述根据设置的所述第一图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数、宽度参数以及第一电流的值，以获得调整后的降维参数γ、宽度参数ω以及第一电流I具体包括：

分别调整所述降维参数、所述宽度参数或所述第一电流的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，每个调整后的图像识别成功率对应于每

个调整后的系统功耗;

确定调整后的每个图像识别成功率与所述第一图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗;

选择在满足所述最小系统功耗时获得最大图像识别成功率的降维参数、宽度参数以及第一电流的值分别作为调整后的所述降维参数γ、所述宽度参数ω以及所述第一电流I。

22、根据权利要求19所述的图像识别方法，其特征在于，还包括:

确定统计的图像识别成功率与设置的第二图像识别成功率的差值的绝对值大于预设阈值;

根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，以获得调整后的降维参数γ'、宽度参数ω'以及第一电流I'，其中，所述第二图像识别成功率与所述第一图像识别成功率不同;

根据调整后的所述降维参数γ'降低第二图像数据的维度;

将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述ω'为调整后的宽度参数，所述 I'小于所述$I_s$;

确定所述NVM中存储的图像库中是否包含有与所述降维后的第二图像数据相匹配的图像数据。

23、根据权利要求19-22任意一项所述的图像识别方法，其特征在于，还包括:

统计在预设的统计期间内所述图像匹配模块输出的匹配结果，获取所述统计的图像识别成功率。

24、根据权利要求19-23任意一项所述的图像识别方法，其特征在于，所述根据设置的降维参数γ降低第一图像数据的维度包括:

根据所述第一图像数据与设置的二进制矩阵的乘积获得所述降维后的第一图像数据，其中，所述第一图像数据为k行*m列的矩阵，所述二进制矩阵为m行*n列的矩阵，所述降维后的第一图像数据为k行*n列的矩阵，k、m 和n 为正整数，m的值大于n，n的

值根据设置的降维参数γ确定，γ=n/m。

25、根据权利要求20-24任意一项所述的图像识别方法，其特征在于，所述根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I，包括：

分别调整降维参数γ、所述宽度参数ω或第一电流I的值，并分别获得多个调整后的图像识别成功率和多个调整后的系统功耗，其中，所述E的值与$\gamma((N-\omega)* I_s^2 +\omega*I)$的值成正比；

确定调整后的每个图像识别成功率与所述第二图像识别成功率之间的差值，选择所述差值的绝对值不大于所述预设阈值时的至少一个调整后的图像识别成功率所对应的至少一个调整后的系统功耗中的最小系统功耗E'；

选择在满足所述最小功耗E'时获得最大图像识别成功率的降维参数、宽度参数以及第一电流的值分别作为所述调整后的降维参数γ'、宽度参数ω'以及第一电流I'。

26、根据权利要求24所述的图像识别方法，其特征在于：所述二进制类型的矩阵包括伯努利映射矩阵。

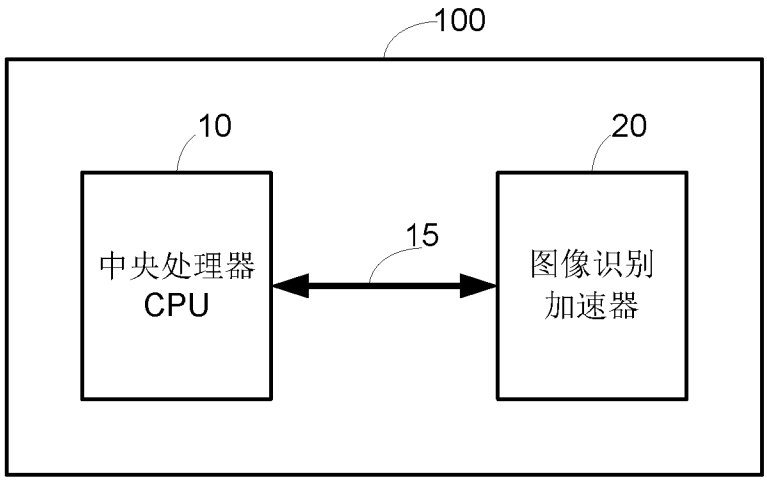
100
10
20
15
中央处理器
CPU
图像识别
加速器

图 1

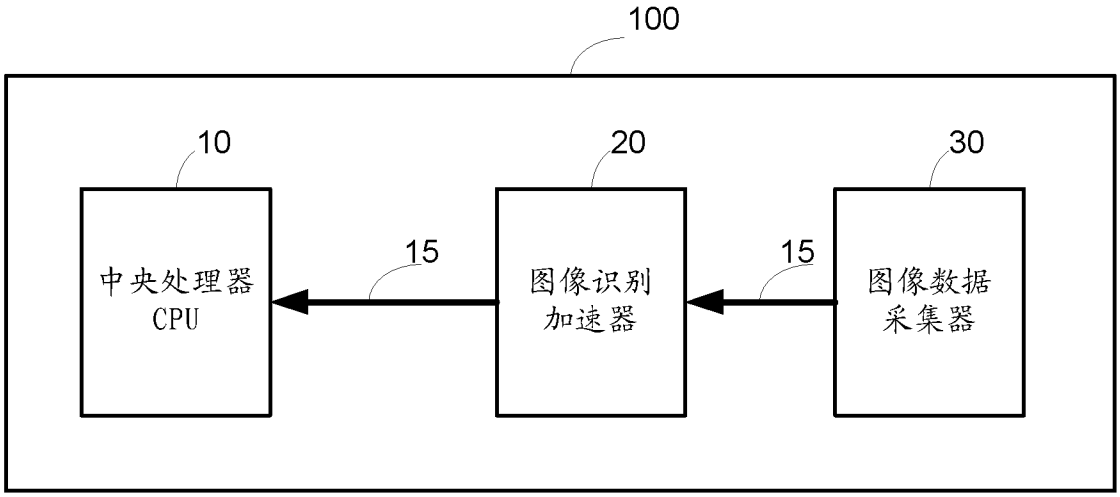
100
10
20
30
15
15
中央处理器
CPU
图像识别
加速器
图像数据
采集器

图 2

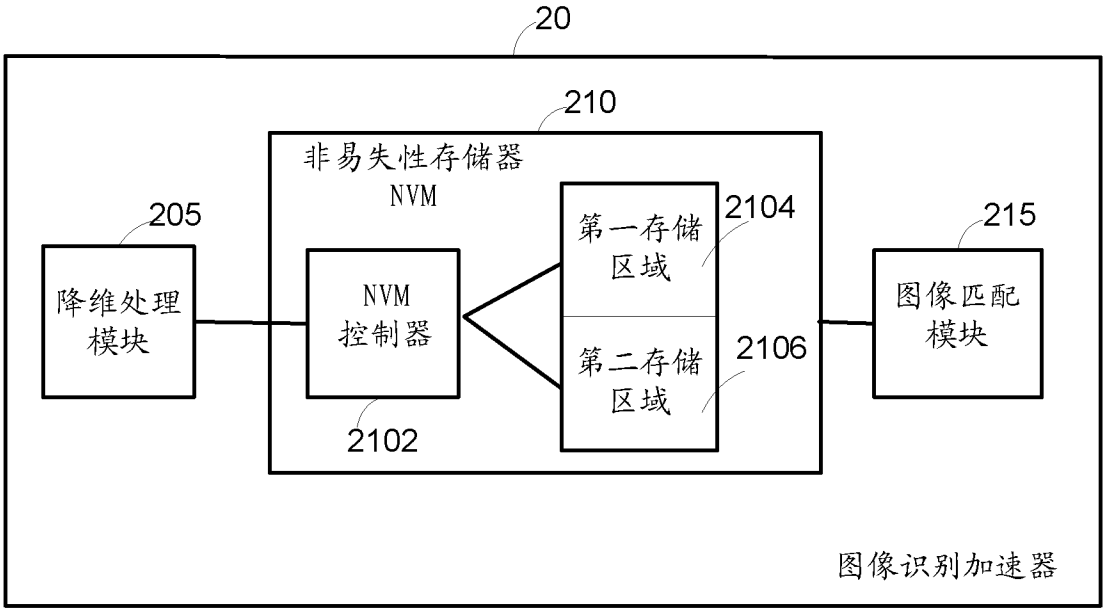
20
210
非易失性存储器
NVM
205
降维处理
模块
NVM
控制器
2102
第一存储
区域
2104
第二存储
区域
2106
215
图像匹配
模块
图像识别加速器

图 3

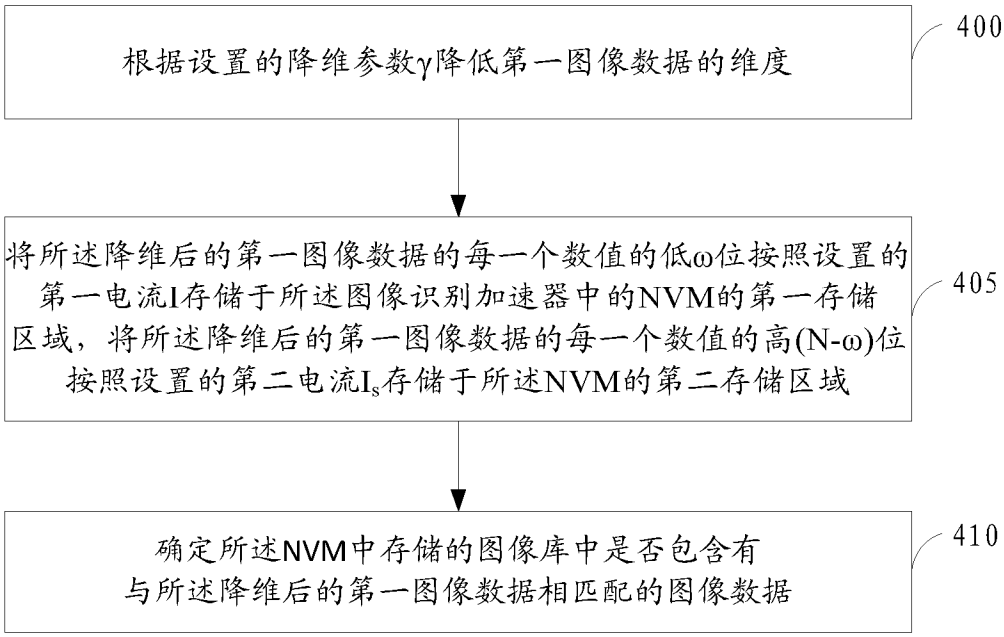
根据设置的降维参数γ降低第一图像数据的维度
400
将所述降维后的第一图像数据的每一个数值的低ω位按照设置的第一电流I存储于所述图像识别加速器中的NVM的第一存储区域，将所述降维后的第一图像数据的每一个数值的高(N-ω)位按照设置的第二电流$I_s$存储于所述NVM的第二存储区域
405
确定所述NVM中存储的图像库中是否包含与所述降维后的第一图像数据相匹配的图像数据
410

图 4

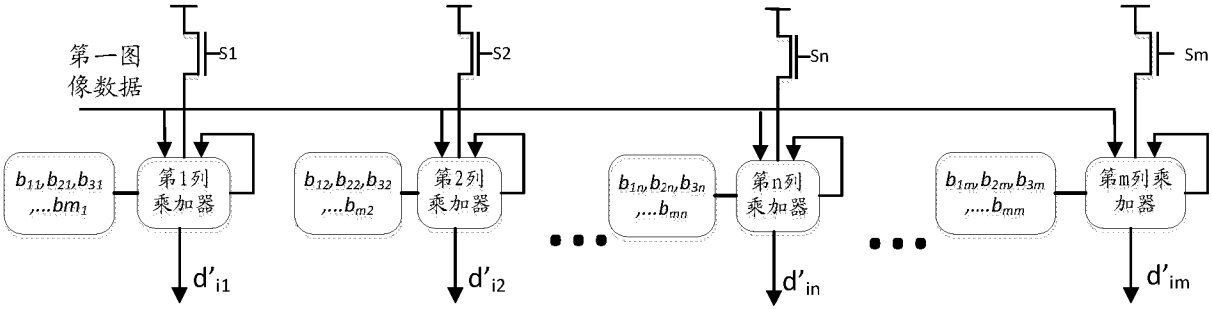
第一图像数据
S1
S2
Sn
Sm
$b_{11}, b_{21}, b_{31}$ ,...$bm_1$
第1列乘加器
$b_{12}, b_{22}, b_{32}$ ,...$b_{m2}$
第2列乘加器
$b_{1n}, b_{2n}, b_{3n}$ ,...$b_{mn}$
第n列乘加器
$b_{1m}, b_{2m}, b_{3m}$ ,....$b_{mm}$
第m列乘加器
$d'_{i1}$
$d'_{i2}$
$d'_{in}$
$d'_{im}$

图 5

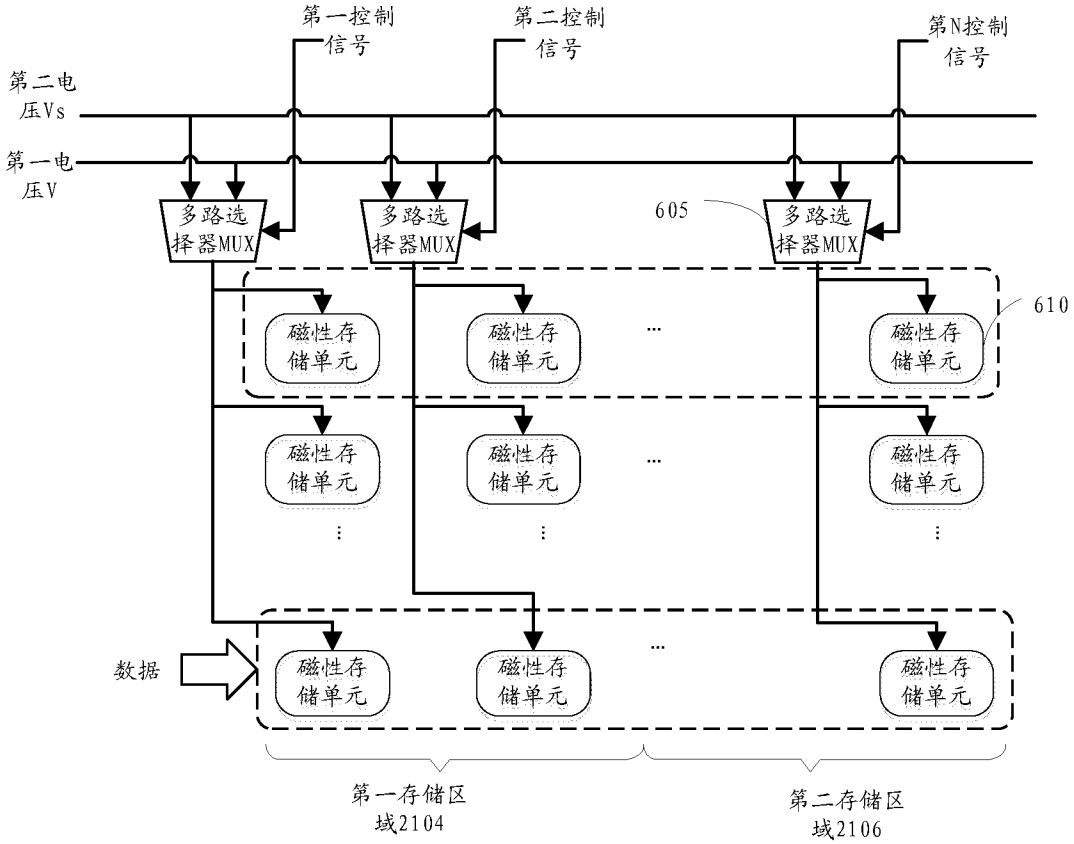
第一控制信号
第二控制信号
第N控制信号
第二电压Vs
第一电压V
多路选择器MUX
多路选择器MUX
605
多路选择器MUX
610
磁性存储单元
数据
第一存储区域2104
第二存储区域2106

图 6

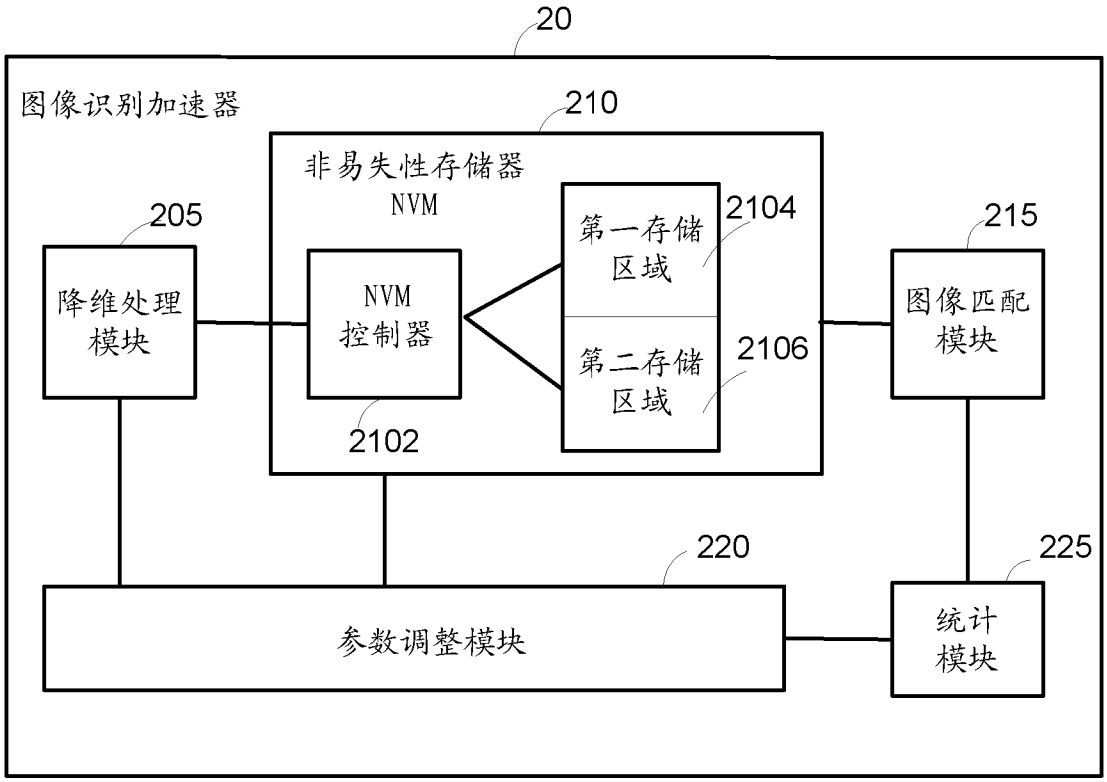
20
图像识别加速器
210
非易失性存储器
NVM
205
降维处理
模块
NVM
控制器
2102
第一存储
区域
2104
第二存储
区域
2106
215
图像匹配
模块
220
参数调整模块
225
统计
模块

图 7

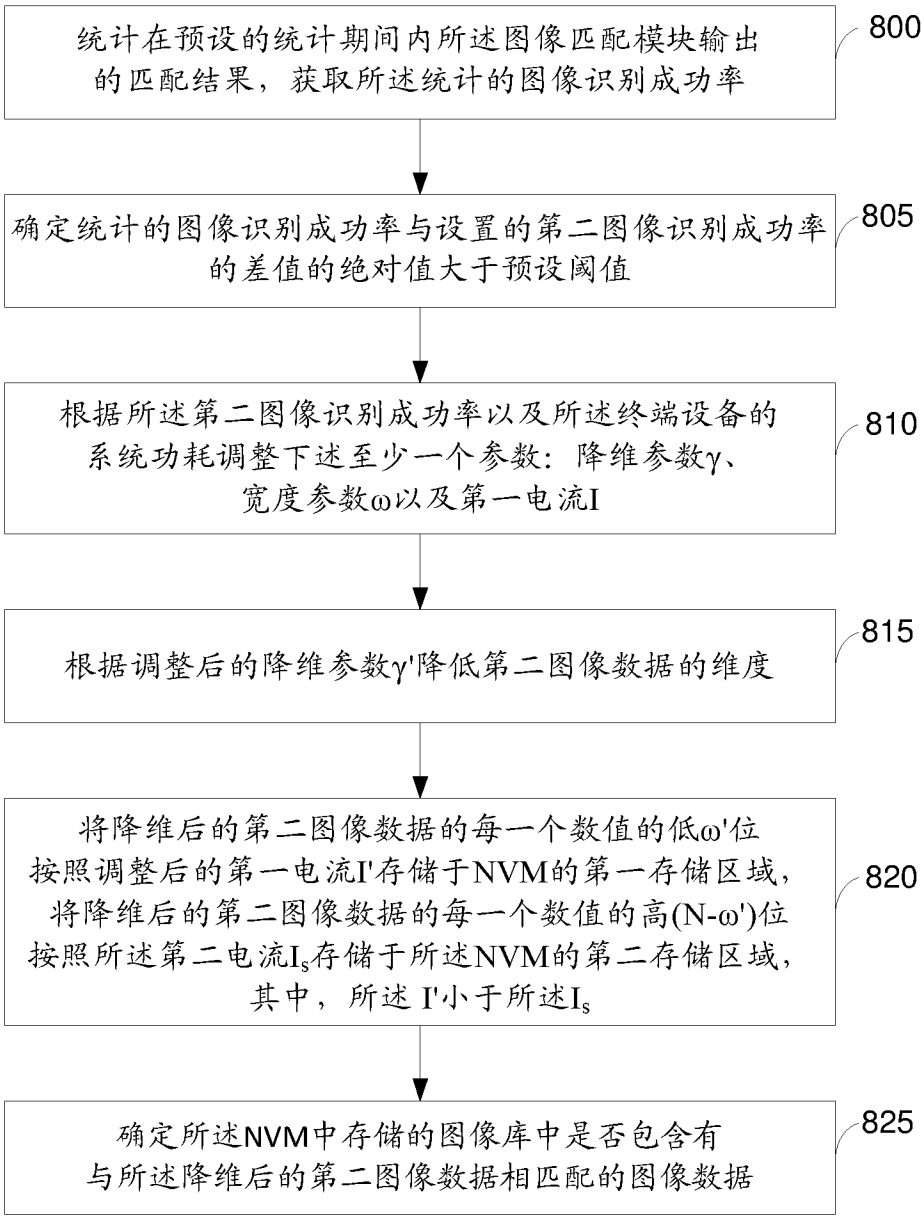
统计在预设的统计期间内所述图像匹配模块输出的匹配结果，获取所述统计的图像识别成功率
800
确定统计的图像识别成功率与设置的第二图像识别成功率的差值的绝对值大于预设阈值
805
根据所述第二图像识别成功率以及所述终端设备的系统功耗调整下述至少一个参数：降维参数γ、宽度参数ω以及第一电流I
810
根据调整后的降维参数γ'降低第二图像数据的维度
815
将降维后的第二图像数据的每一个数值的低ω'位按照调整后的第一电流I'存储于NVM的第一存储区域，将降维后的第二图像数据的每一个数值的高(N-ω')位按照所述第二电流$I_s$存储于所述NVM的第二存储区域，其中，所述 I'小于所述$I_s$
820
确定所述NVM中存储的图像库中是否包含与所述降维后的第二图像数据相匹配的图像数据
825

图 8

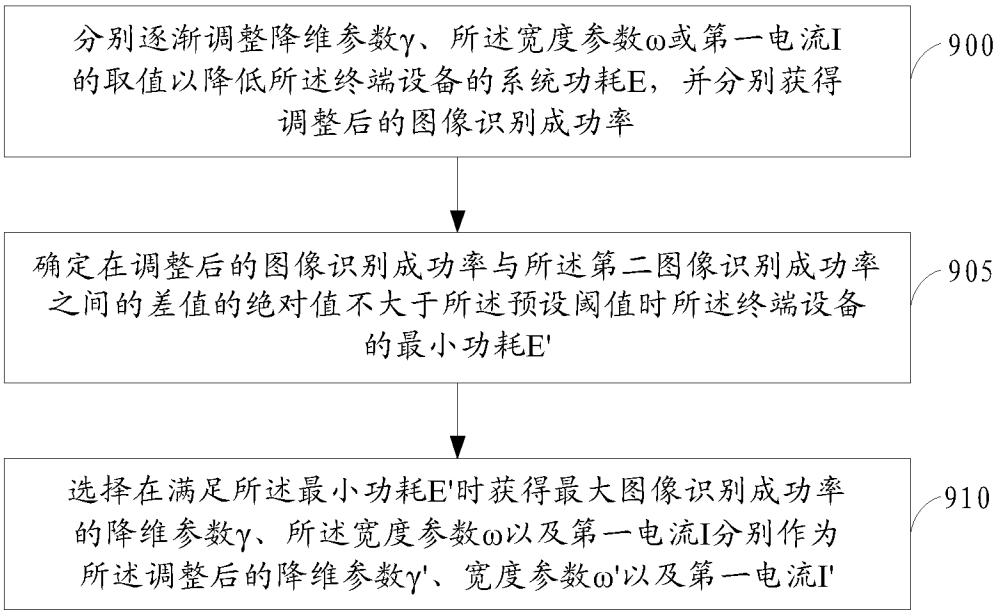


图 9

| 降维参数 γ | 宽度参数 ω | 第一电流 I | 系统功耗 E | 图像识别成功率 Qos |
|---|---|---|---|---|
| γ1 | ω | I | E1 | Qos1 |
| γ2 | ω | I | E2 | Qos2 |
| γ | ω1 | I | E3 | Qos3 |
| γ | ω | I2 | E4 | Qos4 |
| γ3 | ω3 | I3 | E5 | Qos5 |
| … | … | … | … | … |

图 10 (a)

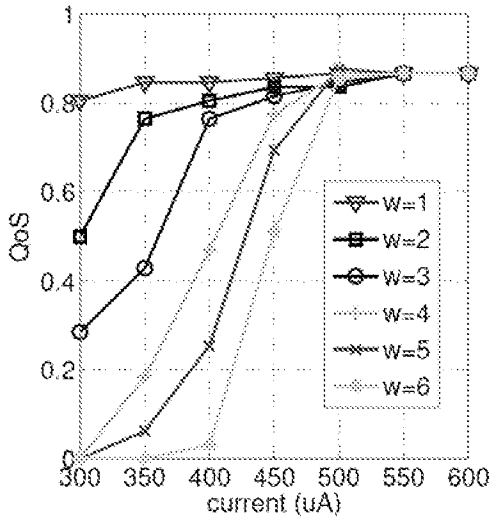
1
0.8
0.6
0.4
0.2
0
QoS
300 350 400 450 500 550 600
current (uA)
W=1
W=2
W=3
W=4
W=5
W=6

图 10 (b)

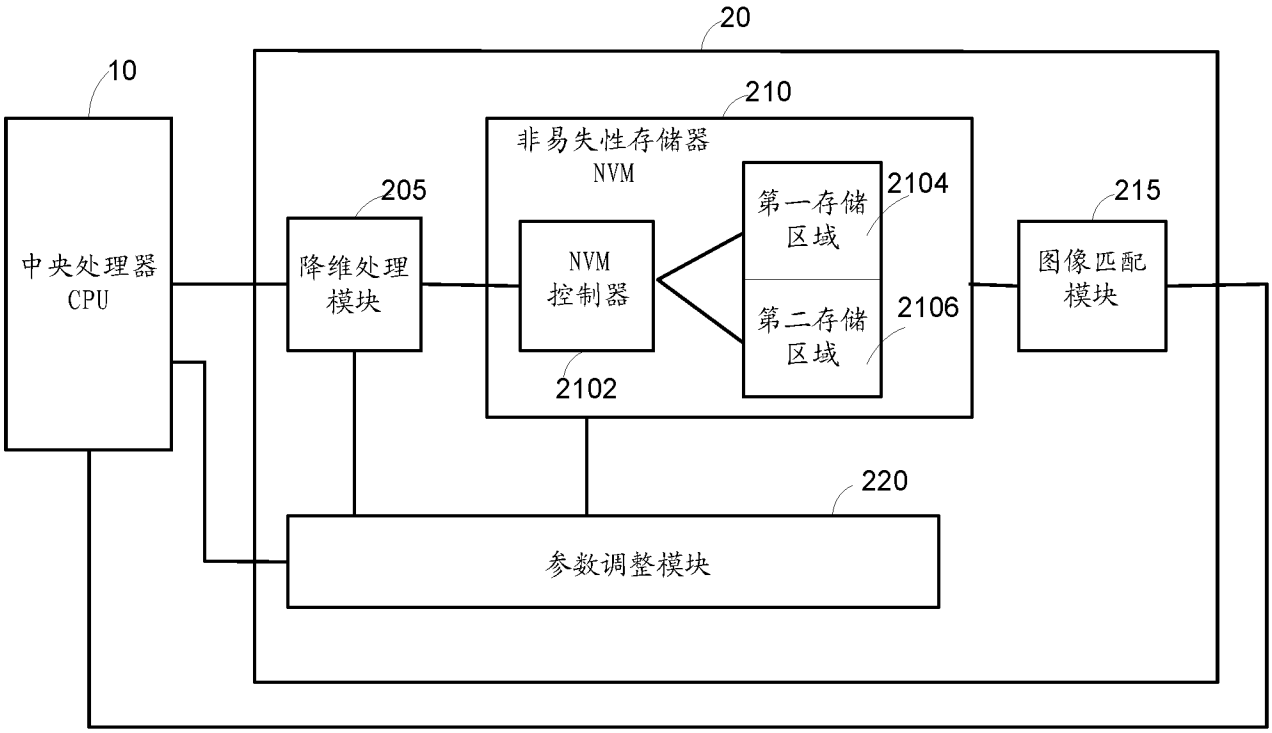
20
10
210
非易失性存储器
NVM
205
2104
215
中央处理器
CPU
降维处理
模块
NVM
控制器
第一存储
区域
第二存储
区域
2106
图像匹配
模块
2102
220
参数调整模块

图 11

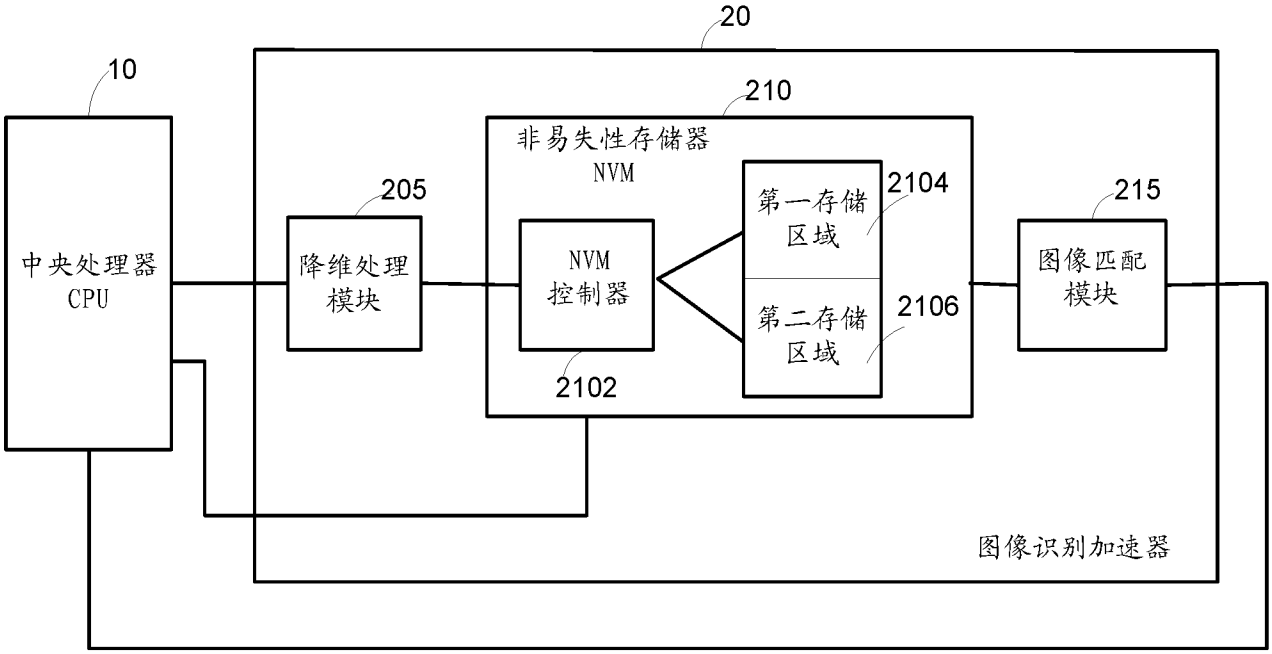
20
10
210
非易失性存储器
NVM
205
2104
215
中央处理器
CPU
降维处理
模块
NVM
控制器
第一存储
区域
图像匹配
模块
2106
第二存储
区域
2102
图像识别加速器

图 12

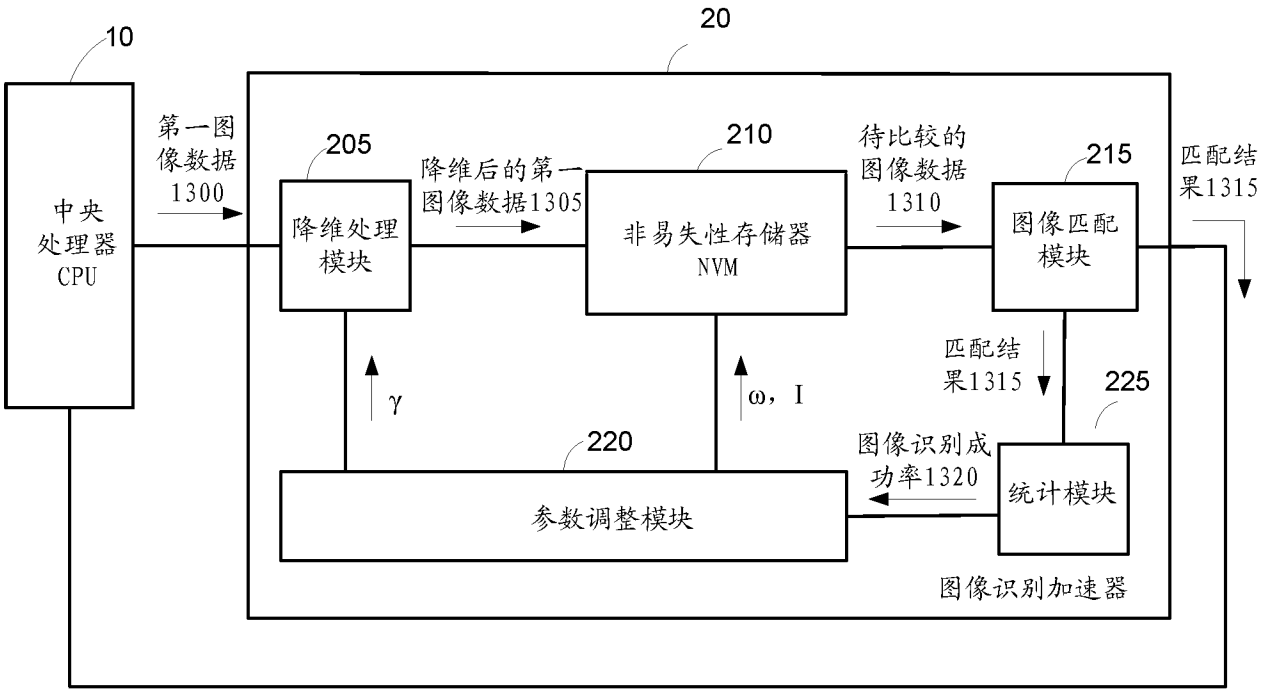
20
10
第一图
像数据
1300
205
降维后的第一
图像数据1305
210
待比较的
图像数据
1310
215
匹配结
果1315
中央
处理器
CPU
降维处理
模块
非易失性存储器
NVM
图像匹配
模块
匹配结
果1315
225
γ
ω，I
220
图像识别成
功率1320
参数调整模块
统计模块
图像识别加速器

图 13

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/CN2016/074240**

**A. CLASSIFICATION OF SUBJECT MATTER**

G06K 9/00 (2006.01) 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)

G06K; G06F; G06T

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)

CNPAT, CNKI, WPI, EPODOC, GOOGLE: reduce, decrease, dimension, dimension reduction, storage, image, identify, match, NVM, non-volatile memory, RRAM, MR AM, FRAM, PCM, current, power consumption

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | CN 1904907 A (SHANGHAI BWAVE TECHNOLOGY CO., LTD.), 31 January 2007 (31.01.2007), description, page 3, paragraph 2 to page 4, paragraph 2, page 7, last paragraph, and figure 5 | 1-26 |
| A | CN 1577622 A (SEMICONDUCTOR ENERGY LABORATORY CO., LTD.), 09 February 2005 (09.02.2005), the whole document | 1-26 |
| A | CN 103810119 A (BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS), 21 May 2014 (21.05.2014), the whole document | 1-26 |
| A | CN 103514432 A (NOKIA CORPORATION), 15 January 2014 (15.01.2014), the whole document | 1-26 |

☐ 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 | Date of mailing of the international search report |
|---|---|
| 21 April 2016 (21.04.2016) | **17 May 2016 (17.05.2016)** |
| Name and mailing address of the ISA/CN:<br>State Intellectual Property Office of the P. R. China<br>No. 6, Xitucheng Road, Jimenqiao<br>Haidian District, Beijing 100088, China<br>Facsimile No.: (86-10) 62019451 | Authorized officer<br>**SUN, Guohui**<br>Telephone No.: (86-10) **62413599** |

# INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
**PCT/CN2016/074240**

| Patent Documents referred in the Report | Publication Date | Patent Family | Publication Date |
|---|---|---|---|
| CN 1904907 A | 31 January 2007 | None | |
| CN 1577622 A | 09 February 2005 | KR 20050009692 A | 25 January 2005 |
| | | DE 602004018432 D1 | 29 January 2009 |
| | | JP 2005038557 A | 10 February 2005 |
| | | TW I354285 B | 11 December 2011 |
| | | US 2005013180 A1 | 20 January 2005 |
| | | EP 1501097 B1 | 17 December 2008 |
| CN 103810119 A | 21 May 2014 | None | |
| CN 103514432 A | 15 January 2014 | US 2015205997 A1 | 23 July 2015 |
| | | WO 2014001610 A1 | 03 January 2014 |
| | | EP 2864933 A1 | 29 April 2015 |

# 国际检索报告

国际申请号
PCT/CN2016/074240

A. 主题的分类

G06K 9/00(2006.01)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)

G06K; G06F; G06T

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））

CNPAT,CNKI,WPI,EPODOC,GOOGLE:图像，识别，匹配，降低，减小，维度，降维，非易失，存储器，存贮器，内存，电流，功耗，image，identify，match，NVM，non-volatile memory，RRAM，MRAM，FRAM，PCM，current，power consumption

C. 相关文件

| 类 型* | 引用文件，必要时，指明相关段落 | 相关的权利要求 |
|---|---|---|
| A | CN 1904907 A (上海明波通信技术有限公司) 2007年 1月 31日 (2007-01-31)<br>说明书第3页第2段-第4页第2段，第7页最后一段，图5 | 1-26 |
| A | CN 1577622 A (株式会社半导体能源研究所) 2005年 2月 9日 (2005-02-09)<br>全文 | 1-26 |
| A | CN 103810119 A (北京航空航天大学) 2014年 5月 21日 (2014-05-21)<br>全文 | 1-26 |
| A | CN 103514432 A (诺基亚公司) 2014年 1月 15日 (2014-01-15)<br>全文 | 1-26 |

☐ 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 | 国际检索报告邮寄日期 |
|---|---|
| 2016年 4月 21日 | 2016年 5月 17日 |
| ISA/CN的名称和邮寄地址<br>中华人民共和国国家知识产权局(ISA/CN)<br>中国北京市海淀区蓟门桥西土城路6号 100088<br>传真号 (86-10)62019451 | 受权官员<br>孙国辉<br>电话号码 (86-10)62413599 |

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/074240

| 检索报告引用的专利文件 | | | 公布日<br>(年/月/日) | 同族专利 | | | 公布日<br>(年/月/日) |
|---|---|---|---|---|---|---|---|
| CN | 1904907 | A | 2007年 1月 31日 | 无 | | | |
| CN | 1577622 | A | 2005年 2月 9日 | KR | 20050009692 | A | 2005年 1月 25日 |
| | | | | DE | 602004018432 | D1 | 2009年 1月 29日 |
| | | | | JP | 2005038557 | A | 2005年 2月 10日 |
| | | | | TW | I354285 | B | 2011年 12月 11日 |
| | | | | US | 2005013180 | A1 | 2005年 1月 20日 |
| | | | | EP | 1501097 | B1 | 2008年 12月 17日 |
| CN | 103810119 | A | 2014年 5月 21日 | 无 | | | |
| CN | 103514432 | A | 2014年 1月 15日 | US | 2015205997 | A1 | 2015年 7月 23日 |
| | | | | WO | 2014001610 | A1 | 2014年 1月 3日 |
| | | | | EP | 2864933 | A1 | 2015年 4月 29日 |