

(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
2017 年 4 月 27 日 (27.04.2017)



(10) 国际公布号
WO 2017/067398 A1

- (51) 国际专利分类号:
G06F 21/36 (2013.01)
- (21) 国际申请号: PCT/CN2016/101731
- (22) 国际申请日: 2016 年 10 月 11 日 (11.10.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510683529.2 2015 年 10 月 20 日 (20.10.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 王曜 (WANG, Yao); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

- (74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, DJ, 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, KW, 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, 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,

[见续页]

(54) Title: METHOD AND DEVICE FOR IDENTIFYING PICTURES BASED ON FILLING IN VERIFICATION CODES BY USER

(54) 发明名称: 一种基于用户填写验证码来进行图片识别的方法及装置

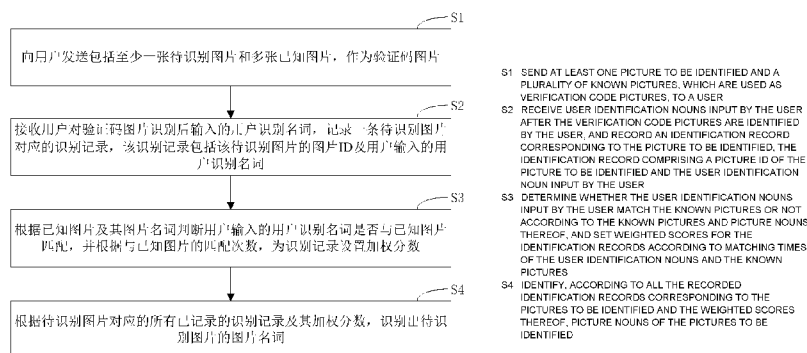


图 1

(57) Abstract: A method and device for identifying pictures based on filling in verification codes by a user. The method comprises: sending at least one picture to be identified and a plurality of known pictures, which are used as verification code pictures, to a user (S1); receiving user identification nouns input by the user after the verification code pictures are identified by the user, and recording an identification record corresponding to the picture to be identified, the identification record comprising a picture ID of the picture to be identified and the user identification noun input by the user (S2); determining whether the user identification nouns input by the user match the known pictures or not according to the known pictures and picture nouns thereof, and setting weighted scores for the identification records according to matching times of the user identification nouns and the known pictures (S3); and finally, identifying, according to all the recorded identification records corresponding to the pictures to be identified and the weighted scores thereof, picture nouns of the pictures to be identified (S4). The device comprises a verification code picture sending module, a receiving module, a weighting module, and an identification module. By means of the method and the device, picture identification processing speed is accelerated, experience in machine learning is increased, and the identification processing capacity for mass pictures is effectively improved.

(57) 摘要:

[见续页]



WO 2017/067398 A1



RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, **本国际公布:**

CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, — 包括国际检索报告(条约第 21 条(3))。
TG)。

一种基于用户填写验证码来进行图片识别的方法及装置，该方法包括：向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片（S1）；接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词（S2）；根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数（S3）；最后根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词（S4）。所述装置包括验证码图片发送模块、接收模块、加权模块和识别模块。所述方法及装置，加快了图片识别处理速度，增加了机器学习经验，有效提高对海量图片的识别处理能力。

一种基于用户填写验证码来进行图片识别的方法及装置

本申请要求 2015 年 10 月 20 日递交的申请号为 201510683529.2、发明名称为“一种基于用户填写验证码来进行图片识别的方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5

技术领域

本发明属于计算机技术领域，尤其涉及一种基于用户填写验证码来进行图片识别的方法及装置。

10 背景技术

互联网的发展始终在“创新与改变”中跨越前行，从联系平台到浏览平台，到交互平台，到工作平台，互联网正逐步深入人们的生活。随着互联网的不断发展，逐步进入大数据时代，城市数据、企业数据、医疗数据、网站数据成为我们虚拟与现实生活的重要组成部分。

15 互联网应用的后台服务器在对大数据进行机器学习的过程中，往往会遇到机器无法准确识别图片的情况，需要大量人力进行标注和纠正。而对于海量的互联网数据，对无法机器学习的图片进行人工处理，是极其繁重和枯燥的工作。

发明内容

20 本发明的目的是提供一种基于用户填写验证码来进行图片识别的方法，将待识别图片作为验证码图片，通过用户身份验证过程中对该验证码图片的识别，来确定该待识别图片的图片名词，巧妙借助用户的识别，对大量待识别图片进行了识别，节省了人力物力。

为了实现上述目的，本发明技术方案如下：

25 一种基于用户填写验证码来进行图片识别的方法，所述方法包括：

向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片；

接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词；

根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并

30 根据与已知图片的匹配次数，为识别记录设置加权分数；

根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

进一步地，所述接收用户对验证码图片识别后输入的用户识别名词之后，还包括：
记录用户本次登录成功与否的结果；

5 所述方法还包括步骤：

根据用户连续登录成功的次数，为识别记录设置额外加权分数；
所述识别记录的加权分数还包括所述额外加权分数。

进一步地，所述根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词之前，还包括：

10 删除加权分数小于等于设定阈值的识别记录。

进一步地，所述方法还包括步骤：

根据用户历史信用评分，对所述加权分数进行修正，修正公式如下：

$$F=\text{round}((F_0+Y+Z)/3, 2)$$

其中 F 为修正后的加权分数， F_0 为修正前的加权分数，Y 为用户历史信用评分，Z
15 为基础分。

本发明所述根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词，包括：

获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待识
20 别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交
25 给人工识别。

进一步地，所述待识别图片对应应为其设置的显示次数，所述向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片，还包括步骤：

验证所述待识别图片提交作为验证码图片的次数是否达到所述显示次数，当达到所述显示次数时，不再提交该图片作为验证码图片。

30 本发明还提出了一种基于用户填写验证码来进行图片识别的装置，所述装置包括：

验证码图片发送模块，用于向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片；

接收模块，用于接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词；

加权模块，用于根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数；

识别模块，用于根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

10 进一步地，所述接收模块还用于记录用户本次登录成功与否的结果；

则，所述加权模块还根据用户连续登录成功的次数，为识别记录设置额外加权分数；所述识别记录的加权分数还包括所述额外加权分数。

进一步地，所述加权模块还用于删除加权分数小于等于设定阈值的识别记录。

进一步地，所述加权模块还用于根据用户历史信用评分，对所述加权分数进行修正，
15 修正公式如下：

$$F=\text{round}((F_0+Y+Z)/3, 2)$$

其中 F 为修正后的加权分数， F_0 为修正前的加权分数，Y 为用户历史信用评分，Z 为基础分。

进一步地，所述识别模块根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词，执行如下操作：

获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待识别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交给人工识别。

进一步地，所述待识别图片对应有其设置的显示次数，所述验证码图片发送模块
30 还用于验证所述待识别图片提交作为验证码图片的次数是否达到所述显示次数，当达到

所述显示次数时，不再提交该图片作为验证码图片。

本发明提出的一种基于用户填写验证码来进行图片识别的方法及装置，将待识别图片作为验证码图片，通过用户身份验证过程中对该验证码图片的识别，来确定该待识别图片的图片名词，巧妙借助用户的识别，对大量待识别图片进行了识别，节省了人力物力。通过本发明的方法和装置，加快了图片识别处理速度，增加了机器学习经验，有效提高对海量图片的识别处理能力。

附图说明

图 1 为本发明基于用户填写验证码来进行图片识别的方法流程图；

图 2 为本发明基于用户填写验证码来进行图片识别的装置结构示意图。

具体实施方式

下面结合附图和实施例对本发明技术方案做进一步详细说明，以下实施例不构成对本发明的限定。

本发明的总体思路是利用用户登录时的验证过程中，需要输入验证码的这一环节，将待识别的图片夹杂在验证图片中发送给用户，用户需要选择图片进行验证，就可以根据用户对待识别的图片的选择来进行分析，识别出待识别的图片。

本实施例为此设置了四个数据表：数据表 A，数据表 B，数据表 C，和数据表 D，其中：

数据表 A 用来存储机器无法识别的图片，也是需要通过本发明进行识别的待识别图片；

数据表 B 用来存储已知的图片及对应的名称；

数据表 C 用来存储识别记录；

数据表 D 用来存储对待识别图片的最终识别结果。

本实施例在用户登录时，向用户发送 4 张图片作为验证码，其中 3 张来自数据表 B，1 张来自数据表 A。需要说明的是，用于作为验证码的图片的具体数量不限，但是仅包括一张待识别图片，本实施例以 4 张为例来进行说明。

要求用户对图片进行识别，返回 4 张图片的用户识别名词。接收到用户返回的用户识别名词后，验证用户的答案，如果用户填对了三张来自数据表 B 的图片名词，就认为其可以通过验证。

基于上述验证过程，本实施例一种基于用户填写验证码来进行图片识别的方法，如图 1 所示，包括：

步骤 S1、向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片。

当用户登录时，需要对其进行身份验证，身份验证通过图片验证码的形式，则需要
5 向用户发送作为验证码的图片。例如假设验证码图片的总数为 4 张，则从数据表 B 中取三个已知图片，以及从数据表 A 中取一张待识别图片，作为验证码图片发送给用户。

本发明所要做的事情就是识别出该待识别图片对应的图片名词，所述图片名词可以是图片中包含的物体的名称。本实施例以 4 张图片作为验证码图片为例，在实际使用中可以提供更多的验证码图片，其中包括至少一张待识别图片以及多张已知的图片，并根据
10 已知图片的识别结果进行身份验证，所包括的待识别图片的多少对身份验证不影响。

步骤 S2、接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID（identification，身份标识）及用户输入的用户识别名词。

本实施例要求用户对验证码图片进行识别，返回 4 张图片的用户识别名词。用户收到验证码图片后，一一输入各个图片的用户识别名词，本实施例将用户输入的图片名词称为用户识别名词。
15

在接收到返回的 4 张图片的用户识别名词后，与已知的图片名词进行对比，如果用户填对了三张来自数据表 B 的已知图片的图片名词时，就认为其可以通过验证。

例如从数据表 B 中选取的三张已知图片分别为电风扇、电视机和洗衣机，而从数据表 A 中选取的一张图片为待识别图片，四张图片打乱次序发送给用户作为验证码图片。
20 假设电风扇排在第一，待识别图片排在第二，电视机排在第三，洗衣机排在第四。

则根据用户返回的用户识别名词，其中对应三张已知图片的图片名词，用户返回的用户识别名词都与已知的图片名词一致，即第一张图片返回的是电风扇，第三张图片返回的是电视机，第四张图片返回的是洗衣机，则判断验证通过。显然待识别图片不会影响判断结果，但是用户还是会返回对应的用户识别名词，能够根据用户返回的用户识别名词进行分析完成对待识别图片的识别，给出其识别的图片名词。
25

然而虽然用户给出了对于待识别图片的用户识别名词，也并不能就此确定该用户给出的用户识别名词就是该待识别图片的真正图片名词，为此需要进一步进行分析。

为了便于对用户返回的对应待识别图片的用户识别名词进行分析，本实施例记录一条识别记录到数据表 C，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别
30

名词。

本实施例还记录用户本次登录成功与否的结果，以便于后续步骤进行加权计算。

步骤 S3、根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数。

- 5 为了对用户返回的对应待识别图片的用户识别名词进行分析，本实施例根据已知图片判断用户输入的用户识别名词正确数量，为识别记录设置加权分数，即加权分数与用户正确地返回已知图片的用户识别名词有关。

例如：本实施例有三张已知图片，按下表为识别记录设置加权 f_1 ：

正确数量	加权分数
0	-1
1	1
2	2
3	5

表 1

- 10 表 1 是计算加权分数的一种数据示意表。

即三张都识别错误，加权分数为-1；一张识别正确，加权分数为 1；两张识别正确，加权分数为 2；三张都识别正确，加权分数为 5。可见如果用户验证通过，则本次验证对应的识别记录的加权分数为 5。在每次验证时，不论是否验证通过，都记录一条识别记录到数据表 C。

- 15 容易理解的是，对应每条识别记录的加权分数，可以以表 1 得到的加权分数为识别记录的加权分数，还可以进一步加入其他因数，如用户历史信用、用户连续登录成功次数等，来进一步提高识别记录加权分数的可信度。

例如，根据记录的用户本次登录成功与否的结果，可以得知用户连续登录成功的次数，则可以根据用户连续登录成功的次数，为识别记录设置额外加权分数。具体的额外

- 20 加权分数 f_2 可以根据下表来计算：

连续登录次数	额外加权分数
连续 5 次失败	-1
没有连续成功	2
连续十次	5

表 2

表 2 是计算加权分数的另一种数据示意表。

至此，当考虑用户连续登录次数时，识别记录的加权分数还包括该额外加权分数，

计算公式如下：

$$F=f_1+f_2$$

需要说明的是，当识别记录的加权分数 F 小于等于设定的阈值（例如-1）时，该识别记录已经不可信，不需要再参与后续的计算，因此本实施例的方法还包括：

- 5 筛选掉加权分数 F 小于等于设定阈值的识别记录。

进一步，本实施例还引入用户历史信用评分来对加权分数 F 进行修正，计算公式如下：

$$F=\text{round}((F_0+Y+Z)/3, 2)$$

- 10 其中， Y 为用户历史信用评分， Z 为基础分， F 为修正后的加权分数， F_0 为修正前的加权分数。可见通过对 F_0 、 Y 、 Z 求平均后，经四舍五入取小数点后两位可得修正后的加权分数 F 。采用修正后的加权分数 F 来进行后续步骤的计算，得到的结果更加可信。

步骤 S4、根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

- 15 随着待识别图片被多个用户识别后，数据表 C 中对应该待识别图片的识别记录越来越多，可以对待识别图片对应的所有已记录的识别记录及其加权分数进行分析，找到待识别图片的准确图片名词。

简单地，可以直接选取已记录的识别记录中加权分数最大的识别记录中的用户识别名词作为该待识别图片的图片名词。

优选地，本实施例采用的方案如下：

- 20 获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待识别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

- 25 根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交给人工识别。

以下进行详细说明：

根据步骤 S1-S3 步骤的计算，不难得到对待识别图片的识别记录及其对应的加权分

数，假设对于一个待识别图片的多条识别记录如下表，表中 F 按照大小顺序排列：

图片 ID	用户识别名词	F_0	Y	F
111	冰箱	F_{0-1}	Y_1	F_1
111	微波炉	F_{0-2}	Y_2	F_2
111	保险箱	F_{0-3}	Y_3	F_3
111	冰箱	F_{0-4}	Y_4	F_4
111	微波炉	F_{0-5}	Y_5	F_5
111	烤箱	F_{0-6}	Y_6	F_6
111	空调	F_{0-7}	Y_7	F_7
111	木柜	F_{0-8}	Y_8	F_8
111	洗衣机	F_{0-9}	Y_9	F_9
111	海尔	F_{0-10}	Y_{10}	F_{10}
...	...	...	...	...

表 3

表 3 是一个示例中待识别图片的多条识别记录示意表。

其中 F_0 为修正前的加权分数， F 为修正后的加权分数， Y 为用户历史评分。

- 5 取该待识别图片 111 的前 M 个 ($M=10$) 条识别记录，将相同的用户识别名词对应的识别记录进行合并，得到如下的合并的识别记录：

图片 ID	用户识别名词	相同用户识别 名词次数 X	加权平均分 数
111	冰箱	2	P_1
111	微波炉	2	P_2
111	保险箱	1	P_3
111	烤箱	1	P_4
111	空调	1	P_5
111	木柜	1	P_6
111	洗衣机	1	P_7
111	海尔	1	P_8

表 4

表 4 是一个示例中合并后的待识别图片的识别记录示意表。

其中，平均加权分数 $P_1 = \text{round}((F_1 + F_3) / 2, 2)$ ， $P_2 = \text{round}((F_2 + F_5) / 2, 2)$ ，即加权平均分数为相同的用户识别名词对应的识别记录的加权分数的平均。

对合并的识别记录按照每条合并的识别记录对应的相同用户识别名词次数进行排序，获取排名第一的合并的识别记录对应的用户识别名词作为待识别图片的图片名词；当有两条以上合并的识别记录的相同用户识别名词次数相同时，取加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词；当两条以上合并的识别记录的相同用户识别名词次数相同，且加权平均分数还是相同，则推送给运维人员，进行人工识别。

最后将识别结果写入数据表 D，作为后续机器学习或者其他业务使用。

需要说明的是，当一个待识别图片经过多次用户识别后，根据本发明的方法已经能确定地进行了识别，则不需要再次将其提交给用户进行识别，因此本发明为每个待识别图片设置了显示次数，当一个待识别图片作为验证码图片提交显示的次数达到该显示次数时，不再提交该待识别图片，将其从数据表 A 中删除。

与上述方法对应地，本实施例一种基于用户填写验证码来进行图片识别的装置如图 2 所述，该装置包括：

验证码图片发送模块，用于向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片；

接收模块，用于接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词；

加权模块，用于根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数；

识别模块，用于根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

与上述方法对应地，接收模块还用于记录用户本次登录成功与否的结果；

则，加权模块还根据用户连续登录成功的次数，为识别记录设置额外加权分数；所述识别记录的加权分数还包括所述额外加权分数。

与上述方法对应地，加权模块还用于删除加权分数小于等于设定阈值的识别记录。

与上述方法对应地，加权模块还用于根据用户历史信用评分，对所述加权分数进行

修正，修正公式如下：

$$F = \text{round}((F_0 + Y + Z) / 3, 2)$$

其中 F 为修正后的加权分数， F_0 为修正前的加权分数，Y 为用户历史信用评分，Z 为基础分。

- 5 与上述方法对应地，识别模块根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词，执行如下操作：

获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待识别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

- 10 根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交给人工识别。

- 15 同样，待识别图片对应应为其设置的显示次数，验证码图片发送模块还用于验证所述待识别图片提交作为验证码图片的次数是否达到所述显示次数，当达到所述显示次数时，不再提交该图片作为验证码图片。

以上实施例仅用以说明本发明的技术方案而非对其进行限制，在不背离本发明精神及其实质的情况下，熟悉本领域的技术人员当可根据本发明作出各种相应的改变和变形，

- 20 但这些相应的改变和变形都应属于本发明所附的权利要求的保护范围。

权利要求书

1、一种基于用户填写验证码来进行图片识别的方法，其特征在于，所述方法包括：
向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片；

5 接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词；

根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数；

根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

10 2、根据权利要求 1 所述的基于用户填写验证码来进行图片识别的方法，其特征在于，所述接收用户对验证码图片识别后输入的用户识别名词之后，还包括：

记录用户本次登录成功与否的结果；

所述方法还包括步骤：

根据用户连续登录成功的次数，为识别记录设置额外加权分数；

15 所述识别记录的加权分数还包括所述额外加权分数。

3、根据权利要求 1 或 2 所述的基于用户填写验证码来进行图片识别的方法，其特征在于，所述根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词之前，还包括：

删除加权分数小于等于设定阈值的识别记录。

20 4、根据权利要求 3 所述的基于用户填写验证码来进行图片识别的方法，其特征在于，所述方法还包括步骤：

根据用户历史信用评分，对所述加权分数进行修正，修正公式如下：

$$F=\text{round}((F_0+Y+Z)/3, 2)$$

25 其中 F 为修正后的加权分数， F_0 为修正前的加权分数，Y 为用户历史信用评分，Z 为基础分。

5、根据权利要求 1 所述的基于用户填写验证码来进行图片识别的方法，其特征在于，所述根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词，包括：

获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相

同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待识别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交给人工识别。

6、根据权利要求 1 所述的基于用户填写验证码来进行图片识别的方法，其特征在于，所述待识别图片对应有为其设置的显示次数，所述向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片，还包括步骤：

验证所述待识别图片提交作为验证码图片的次数是否达到所述显示次数，当达到所述显示次数时，不再提交该图片作为验证码图片。

7、一种基于用户填写验证码来进行图片识别的装置，其特征在于，所述装置包括：验证码图片发送模块，用于向用户发送包括至少一张待识别图片和多张已知图片作为验证码图片；

接收模块，用于接收用户对验证码图片识别后输入的用户识别名词，记录一条待识别图片对应的识别记录，该识别记录包括该待识别图片的图片 ID 及用户输入的用户识别名词；

加权模块，用于根据已知图片及其图片名词判断用户输入的用户识别名词是否与已知图片匹配，并根据与已知图片的匹配次数，为识别记录设置加权分数；

识别模块，用于根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词。

8、根据权利要求 7 所述的基于用户填写验证码来进行图片识别的装置，其特征在于，所述接收模块还用于记录用户本次登录成功与否的结果；

则所述加权模块还根据用户连续登录成功的次数，为识别记录设置额外加权分数；所述识别记录的加权分数还包括所述额外加权分数。

9、根据权利要求 7 或 8 所述的基于用户填写验证码来进行图片识别的装置，其特征在于，所述加权模块还用于删除加权分数小于等于设定阈值的识别记录。

10、根据权利要求 9 所述的基于用户填写验证码来进行图片识别的装置，其特征在于，所述加权模块还用于根据用户历史信用评分，对所述加权分数进行修正，修正公式

如下：

$$F = \text{round}((F_0 + Y + Z) / 3, 2)$$

其中 F 为修正后的加权分数，F₀ 为修正前的加权分数，Y 为用户历史信用评分，Z 为基础分。

- 5 11、根据权利要求 7 所述的基于用户填写验证码来进行图片识别的装置，其特征在于，所述识别模块根据待识别图片对应的所有已记录的识别记录及其加权分数，识别出待识别图片的图片名词，执行如下操作：

获取待识别图片对应的所有识别记录中加权分数排名靠前的 M 条识别记录，将相同用户识别名词的识别记录进行合并，得到合并的识别记录，合并的识别记录包括该待
10 识别图片的图片 ID、对应的用户识别名词、相同用户识别名词次数及加权平均分数；

根据合并的识别记录，筛选出相同用户识别名词次数最大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当有两条以上相同用户识别名词次数最大的合并的识别记录时，以加权平均分数大的合并的识别记录的用户识别名词作为待识别图片的图片名词，当两条以上相同用户识别名词次数最大的合并的加权平均分数相同时，提交
15 给人工识别。

12、根据权利要求 7 所述的基于用户填写验证码来进行图片识别的装置，其特征在于，所述待识别图片对应有为其设置的显示次数，所述验证码图片发送模块还用于验证所述待识别图片提交作为验证码图片的次数是否达到所述显示次数，当达到所述显示次数时，不再提交该图片作为验证码图片。

20

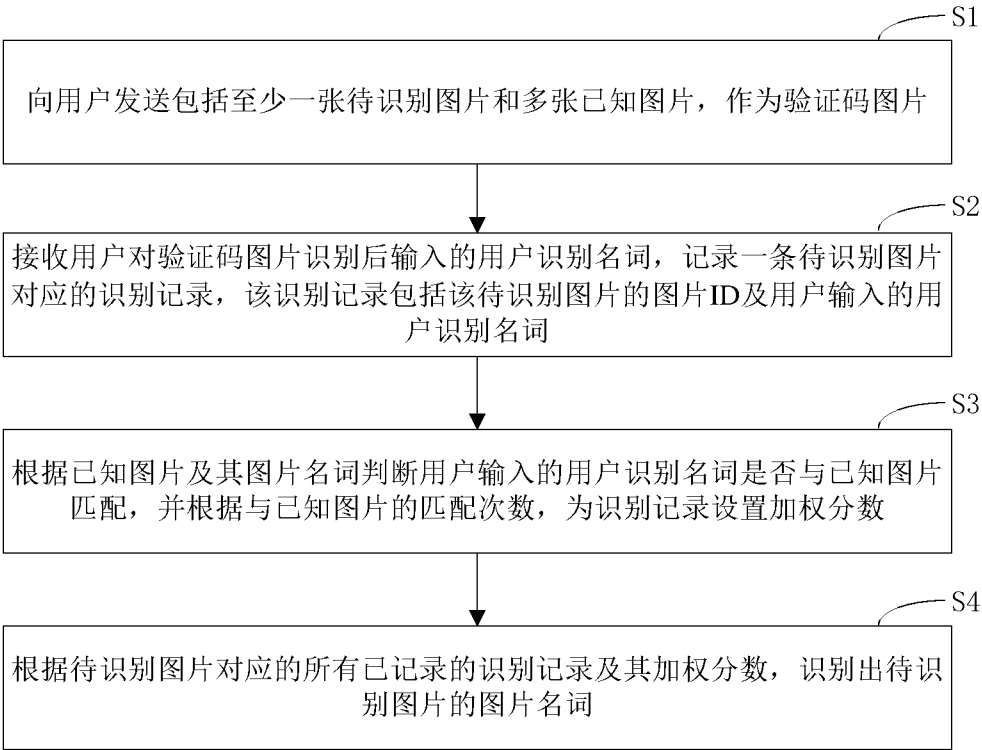


图 1

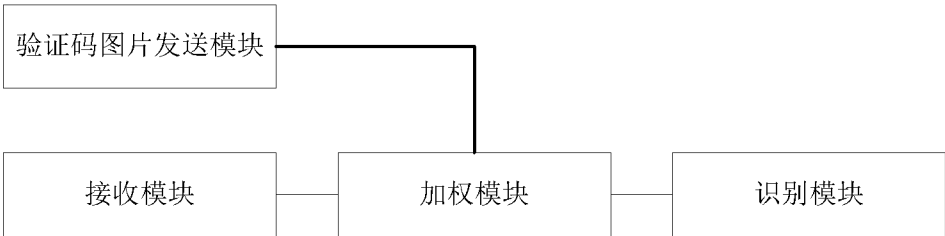


图 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/101731

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/36 (2013.01) 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)

G06F

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: identifying code, image, code, text, identif+, verify+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101388078 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 18 March 2009 (18.03.2009), description, page 5, the last paragraph to page 8, paragraph 1, and figure 4	1-12
X	CN 103186781 A (BEIJING FEINNO COMMUNICATION TECH CO., LTD.), 03 July 2013 (03.07.2013), description, paragraphs [0043]-[0045], and figure 4	1-12
X	CN 102467653 A (FOUNDER INTERNATIONAL SOFTWARE CO., LTD. et al.), 23 May 2012 (23.05.2012), description, paragraphs [0044]-[0060], and figure 2	1-12
A	US 2006274369 A1 (YAMAMOTO, Y.), 07 December 2006 (07.12.2006), the whole document	1-12

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 28 December 2016 (28.12.2016)	Date of mailing of the international search report 16 January 2017 (16.01.2017)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer CHENG, Xiaomei Telephone No.: (86-10) 62413598

International application No.
PCT/CN2016/101731

Form PCT/ISA/210 (patent family annex) (July 2009)

A. 主题的分类 G06F 21/36 (2013.01) 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) G06F 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, EPDOC, GOOGLE: 验证码, 图, 文本, 识别, identifying code, image, code, text, identif+, verify+		
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
X	CN 101388078 A (腾讯科技深圳有限公司) 2009年 3月 18日 (2009 - 03 - 18) 说明书第5页最后1段至第8页第1段、图4	1-12
X	CN 103186781 A (北京新媒传信科技有限公司) 2013年 7月 3日 (2013 - 07 - 03) 说明书第[0043]-[0045]段、图4	1-12
X	CN 102467653 A (方正国际软件北京有限公司 等) 2012年 5月 23日 (2012 - 05 - 23) 说明书第[0044]-[0060]段、图2	1-12
A	US 2006274369 A1 (YAMAMOTO, YOHEI) 2006年 12月 7日 (2006 - 12 - 07) 全文	1-12
☐ 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 2016年 12月 28日		国际检索报告邮寄日期 2017年 1月 16日
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10) 62019451		受权官员 程小梅 电话号码 (86-10) 62413598

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/101731

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	101388078	A	2009年 3月 18日	无	
CN	103186781	A	2013年 7月 3日	无	
CN	102467653	A	2012年 5月 23日	无	
US	2006274369	A1	2006年 12月 7日	JP 2007013934 A	2007年 1月 18日