[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G07D 7/20 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200680047516.5

[43] 公开日 2008 年 12 月 24 日

[11] 公开号 CN 101331527A

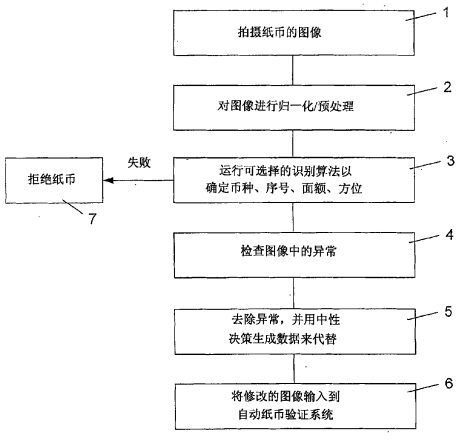
[22] 申请日 2006.12.14
[21] 申请号 200680047516.5
[30] 优先权
[32] 2005.12.16 [33] US [31] 11/305,537
[32] 2006. 3. 2 [33] US [31] 11/366,147
[86] 国际申请 PCT/GB2006/004663 2006.12.14
[87] 国际公布 WO2007/068923 英 2007.6.21
[85] 进入国家阶段日期 2008.6.16
[71] 申请人 NCR 公司
地址 美国俄亥俄州
[72] 发明人 何 超 佳里·罗斯
[74] 专利代理机构 中原信达知识产权代理有限责任公司
代理人 夏 凯 钟 强
权利要求书 4 页 说明书 18 页 附图 7 页

[54] 发明名称

在验证之前处理介质对象的图像

[57] 摘要

在被损坏或标记的介质对象的情况下，自动介质对象验证通常是有问题的。 描述了一种解决该问题的在自动验证之前对介质对象的图像进行处理的方法。 通过例如使用带通滤波器识别异常图像元素。 用中性决策生成数据代替异常图像元素。 该数据关于作为指定自动介质对象验证处理的决策生成处理是中性的。 例如，对于每个异常图像元素，获取图像定位遍及在介质对象的图像训练集合的所有图像中的估计分布。 基于置信度从估计的分布中选择值，所述置信度与自动介质对象验证处理使用的置信度有关。 以这种方式，自动介质对象验证器可成功地处理具有破缝、洞、标记或污迹的介质对象。



1. 一种处理介质对象的图像的方法，包括：

（i）识别所述图像中的异常；

（ii）通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的数据与作为预先指定的介质对象验证处理的决策生成处理有关。

2. 如权利要求1所述的方法，其中，所述识别所述图像中的异常的步骤包括应用带通滤波器。

3. 如权利要求1所述的方法，其中，所述方法包括：对于每个异常图像元素，通过获取图像定位遍及在介质对象的图像的训练集合的所有图像中的估计分布，并从所述估计分布中选择值，获得所述中性决策生成数据。

4. 如权利要求3所述的方法，其中，所述值基于显著水平从所述估计分布中选择，所述显著水平是所述预先指定的介质对象验证处理的显著水平。

5. 如权利要求3所述的方法，其中，所述介质对象的图像的训练集合只包括真介质对象的图像。

6. 如权利要求3所述的方法，其中，所述分布是基于像素亮度简档来估计的。

7. 如权利要求1所述的方法，其中，所述预先指定的介质对象验证处理包括使用单类分类器。

8. 如权利要求1所述的方法，进一步包括：将所述修改的图像作为

输入提供给所述预先指定的介质对象验证处理。

9. 一种处理介质对象的图像的设备，所述设备包括：

（i）处理器，配置以识别所述图像中的异常；

（ii）图像修改器，配置以通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的数据与作为预先指定的介质对象验证处理的决策生成处理有关。

10. 如权利要求9所述的设备，其中，所述处理器包括用于识别所述图像中的异常的带通滤波器。

11. 如权利要求9所述的设备，其中，所述图像修改器被配置用于：对于每个异常图像元素，通过获取图像定位遍及在介质对象的图像的训练集合的所有图像中的估计分布，并从所述估计分布中选择值，获得所述中性决策生成数据。

12. 如权利要求11所述的设备，其中，所述图像修改器被配置用于：基于显著水平从所述估计分布中选择所述值，所述置信度是所述预先指定的介质对象验证处理的置信度。

13. 如权利要求11所述的设备，其中，所述图像修改器被配置用于基于像素亮度简档估计所述分布。

14. 如权利要求11所述的设备，其中，所述配置图像修改器被配置用于：从只包括真介质对象的图像的图像训练集合中估计所述分布。

15. 如权利要求9所述的设备，包括纸币验证器，并且其中，所述配置图像修改器被配置用于：将所述修改的图像输入到所述介质对象验证器。

16. 如权利要求15所述的设备，其中，所述介质对象验证器包括单类分类器。

17. 一种介质对象验证器，包括：

（i）输入，配置以接收将被验证的介质对象的至少一个图像；

（ii）处理器，配置以识别所述图像中的异常；

（iii）图像修改器，配置以通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的数据与所述介质对象验证器的分类器有关；

（iv）分割映射；

（v）处理器，配置以使用所述分割映射分割所述介质对象的图像；

（vi）特征提取器，配置以从所述介质对象的图像的每个分段中提取一个或多个特征；

（vii）分类器，配置以基于提取的特征对所述介质对象进行分类；

其中，所述分割映射包括与介质对象的训练图像集合中所有图像之间的相应图像元素的关系有关的信息。

18. 如权利要求17所述的介质对象验证器，其中，所述图像修改器被配置用于：对于每个异常图像元素，通过获取图像定位遍及在介质对象的图像的训练集合的所有图像中的估计分布，并从所述估计分布中选择值，获得所述中性决策生成数据。

19. 一种包括计算机程序代码装置的计算机程序，所述计算机程序代码装置适用于执行处理纸币的图像的方法的所有步骤，所述方法包括：

当在计算机上运行所述程序时，

（i）识别所述图像中的异常；

（ii）通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的数据与作为预先指定的介质对象验证处理的决策生成处理有关。

20. 如权利要求19所述的计算机程序，所述计算机程序包含在计算机可读介质上。

21. 一种自助式设备，包括：

（i）用于接受介质对象的装置；

（ii）用于获得所述介质对象的数字图像的成像装置；以及

（iii）介质对象验证器，包括：

（i）输入，配置以接收将被验证的介质对象的至少一个图像；

（ii）处理器，配置以识别所述图像中的异常；

（iii）图像修改器，配置以通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的数据与介质对象验证的分类器有关；

（iv）分割映射；

（v）处理器，配置以使用所述分割映射来分割所述介质对象的图像；

（vi）特征提取器，配置以从所述介质对象的图像的每个分段中提取一个或多个特征；

（vii）分类器，配置以基于提取的特征对所述介质对象进行分类；

其中，所述分割映射包括与介质对象的训练图像集合中所有图像之间的相应图像元素的关系有关的信息。

在验证之前处理介质对象的图像

相关申请的交叉引用

本申请是2006年3月2日提交的美国专利申请11/366,147号的部分继续申请，美国专利申请11/366,147号是2005年12月16日提交的美国专利申请11/305,537号的部分继续申请。2006年3月2日提交的申请11/366,147号和2005年12月16日提交的申请11/305,537号通过引用合并于此。

技术领域

本发明涉及一种在验证之前处理介质对象的图像的方法和设备。具体地，涉及处理诸如纸币、护照、债券、股票、支票等的介质对象的图像，但决不限于此。

背景技术

越来越需要以简单、可靠和节约成本的方式自动核查和验证不同币种和面额的纸币。例如，这在接收纸币的自助式设备（例如，自助式售货亭、售票机、被安排办理存款的自动柜员机、自助式货币兑换机等）中是必需的。

以前，货币验证的人工方法涉及纸币的图像检查、诸如水印和纹路对准标记的透视效果以及手感甚至气味。其他已知方法依赖需要半人工询问的半公开特征。例如，使用磁性装置、紫外线传感器、荧光、红外线检测器、电容、金属条、图像模式等。然而，就其本身来说，这些方法是人工或半人工的，并且不适合无法长时间人工干预的许多应用。例如，在自助式设备中。

为创建自动货币验证器有重要问题有待克服。例如，存在具有不

同安全特征甚至基质类型的许多不同类型的货币。在不同的面额还通常具有不同等级的安全特征。因此，需要提供一种用于那些不同币种和面额的、容易并便于执行货币验证的一般方法。

以前的自动验证方法通常需要相对大量的已知假钞样本以训练分类器。另外，那些先前的分类器被训练以只检测已知的假钞。这是有问题的，原因在于，对于可能的假钞存在很少或者不存在可用信息。例如，这特别对于新引入的面额或新引入的币种是有问题的。

在出版在Pattern Recognition 37(模式识别第37期)（2004）第1085-1096页、作者为Chao He, MarkGirolami和Gary Ross（其中二人是本申请的发明人）、题目为“Employing optimized combinations of one-class classifiers for automated currency validation（为自动纸币验证采用单类分类器的最优组合）”的早期论文中，描述了自动纸币验证方法（专利号为EP1484719，US20042447169）。这涉及使用网格结构将整个纸币的图像分割成区域。为每个区域构造单独的“单类（one-class）”分类器，并且组合区域特定分类器的小的子集以提供全面描述（下面更详细地解释术语“单类”）。通过采用遗传算法来实现为实现良好性能的区域特定分类器的分割和组合。该方法在遗传算法阶段中需要少量伪造样本，同样地，当无法获得伪造数据时，该方法不适用。

还需要以能实时执行的计算开销小的方式来执行自动纸币验证。

在纸币被损坏或被做标记的情况下，自动货币验证通常是有问题的。例如，如果纸币有破缝、洞、污迹和/或折叠的角。纸币的老化和在纸币的磨损期间发生的变脏对于自动货币验证系统也是有问题的。

上面提到的许多问题还应用于诸如护照、股票、债券、支票等其他类型的介质的验证。

发明内容

在被损坏或标记的介质对象（media items）的情况下，自动介质对象验证通常是有问题的。描述了一种解决该问题的在自动验证之前对介质对象的图像进行处理的方法。通过例如使用带通滤波器识别异常图像元素。用中性决策生成数据代替异常图像元素。该数据关于作为指定自动介质对象验证处理的决策生成处理是中性的。例如，对于每个异常图像元素，获取图像定位遍及在介质对象的图像训练集合的所有图像中的估计分布。基于置信度从估计的分布中选择值，所述置信度与自动介质对象验证处理使用的置信度有关。以这种方式，自动介质对象验证器可成功地处理具有破缝、洞、标记或污迹的介质对象。

这里所描述的方法可通过在存储介质上的机器可读形式的软件执行。可以以本领域的技术人员清楚的任何合适的顺序和/或并行执行所述方法的步骤。

这表示，软件可以是有价值的、可分开买卖的商品。这意味着包括运行或控制“（哑元）dumb”或标准软件的软件，以执行期望的功能，（因此，软件实质上定义寄存器的功能，因此即使在将其与其标准硬件组合之前，也可被称为寄存器）。由于相似的理由，还包括“描述”或定义硬件的构造的软件，例如设计硅芯片或构造通用可编程芯片所使用的HDL（硬件描述语言）软件，以执行期望的功能。

如技术人员明了的，优选的特征可适当组合，并可与本发明的多方面中的任何特征组合。

附图说明

将参考下面的附图通过示例方式来描述本发明的实施例，其中：

图1是识别和代替纸币图像中的异常图像元素的方法的流程图；

图2是创建用于纸币验证的分类器的方法的流程图；

图3是代替纸币图像中的异常图像元素的方法的流程图；

图4是创建用于纸币验证的分类器的设备的示意图；

图5是纸币验证器的示意图；

图6是验证纸币的方法的流程图；

图7是具有纸币验证器的自助式设备的示意图。

具体实施方式

下面只作为示例，描述本发明的实施例。这些示例表示申请人目前所知道的将本发明应用于实践的最好方式，而不是能实现本发明的仅有方式。尽管这里将本示例描述和示出为用于自动纸币验证被实现，但这里描述的系统作为示例被描述，而不是限制。本领域技术人员应当明白，本示例适合于多种不同类型的介质验证系统中的应用，所述介质验证系统包括护照验证系统、支票验证系统、用于债券和股票的验证系统，但不限于此。

使用术语“单类分类器”来表示使用关于只来自单个类的样本的信息而形成或构造的分类器，但其用于将新出现的样本分配给所述单个类或者不分配给所述单个类。这与传统二进制分类器不同，所述传统二进制分类器是通过使用关于两个类的样本的信息被创建，并被用于将新样本分配给所述两个类中的一个或另一个。单类分类器可被认为定义已知类周围的边界，从而脱离所述边界的样本被认为不属于所述已知类。

如上所述，自动货币验证在被损坏或被标记的纸币的情况下通常是有问题的。例如，如果纸币有破缝、洞、污迹和/或折叠的角。纸币的老化和在纸币的磨损期间发生的变脏对于自动纸币验证系统也是有问题的。

例如，自动纸币验证系统可使用处理，从而将待验证的纸币的图像划分成分段。可通过使用网格结构或单独使用空间位置信息的其他

方法来形成所述分段。可选择地，可通过使用分割映射来形成分段，所述分割映射使用与训练纸币图像的集合的每个项中的相应图像元素之间的图像元素的相关值有关的信息。

如果将被验证的纸币被损坏或被做标记，则因为一些信息异常或不可靠，所以这在自动纸币验证处理中导致问题。例如，纸币中的洞可导致纸币的图像中异常高亮度（intensity）的像素。另外，纸币上的污迹或标记可导致纸币的图像中异常低亮度的像素。

在将待验证的纸币的图像划分成分段作为验证处理的一部分的情况下，一种选择是忽略那些包含异常数据（例如，洞、标记、折痕、破缝等）的分段。然而，只使用少量分段，这意味着大比例的数据被忽略。另外，如果忽略的分段恰巧包含诸如安全特征（例如，全息图、丝纹和水印等）的重要纸币区域，则纸币验证器的置信度将下降。

为了解决这些问题，我们识别诸如将被验证的纸币的介质对象的图像中的异常图像元素，并用决策-中性（decision-neutral）数据来代替那些异常图像元素。通过“决策-中性数据”或“中性决策产生数据”，我们表示将影响预执行介质对象验证处理的结果的数据。所述介质对象验证处理可以是任何合适类型，包括这里描述的特定介质对象验证处理，但不限于此。

图1是处理将被验证的纸币的图像的方法的概要流程图。

使用下面将更详细描述的任何合适的技术来拍摄将被验证的纸币的图像（参看方框1）。将该图像归一化和/或预处理（参看方框2），例如在特定方位对齐所述图像，以及将所述图像缩放到特定大小。这使得传感器和照明环境的变化将被考虑。然后，引入可选步骤（参看方框3）使用识别算法来确定纸币的币种、序号、面额和方位中的一个或多个。如果识别算法失败，则可通过参考纸币图像的不同边缘或角

再次尝试。如果四个边缘都尝试了并且都失败了，则该纸币被拒绝（参看方框7）。否则，处理继续，并寻找图像中的异常（参看方框4）。

可以以任何合适的方式来识别异常。例如，纸币中缺失的区域或洞通常引起异常高亮度的图像区域。在这种情况下，可将指定阈值之上的所有图像区域、元素或像素识别为异常。

在一些币种中，对塑质货币使用窗。这种窗也引起高亮度的图像区域。为了这些窗不被识别为异常，当识别异常时可考虑与这些窗的预期位置、定位和大小有关的知识。

污迹、马克笔标记、钉书钉、折痕和其他这种损坏在纸币图像中引起极度不透明区域。在这种情况下，具有低于指定阈值的亮度的所有图像区域、元素或像素可被识别为异常。可选地，当识别异常时，可考虑与特定币种和面额的图像元素的预期亮度有关的信息。

为了快速识别具有指定阈值之上或之下的亮度的图像元素，可使用带通滤波器。

一旦识别出异常，则可通过用决策中性数据代替来去除异常（参看方框 5）。可选地，对被识别为异常的纸币图像的比例进行检查。如果该比例在指定阈值之上，则如果在拒绝算法阶段还没有拒绝该纸币，则拒绝该纸币（参看方框 7）。这保证由拼合部分模糊假钞的部分真钞而形成的假钞被拒绝。另外，以这种方式可对可能被代替的异常数据的量设置界限。因为处理趋向决策-中性数据代替 100%的纸币图像，所以可能降低检测伪造的能力。

然后，纸币的结果修改图像被传送到纸币验证系统（参看方框 6）以被验证。

下面将参照图 3 更详细地描述形成决策中性数据的处理。

在特定组的实施例中，预先指定的纸币验证处理使用如现在描述所形成的分类器。

图 2 是创建用于纸币验证的分类器的方法的概要流程图。

首先，我们获得真钞的图像的训练集合（参看图 1 的方框 10）。这些是获取的相同币种和面额的纸币的相同类型的图像。图像的类型涉及如何获得图像，这可以用本领域中已知的任何方式。例如，反射图像，透射图像，红、蓝或绿任何通道上的图像，热图像，红外图像，紫外图像，x 光图像或其他图像类型。训练集合中的图像被对准，并且大小相同。如在现有技术中已知，如果必要，可执行预处理来对齐图像并缩放图像。

接着，我们通过使用来自训练集合图像的信息创建分割映射（参看图 2 的方框 12）。所述分割映射包括关于如何将图像划分成多个分段（segment）的信息。所述分段可以是不连续的，即给定分段可包括图像的不同区域中的多于一个的片(patch)。优选地，但非必须，分割映射还包括将被使用的特定数量的分段。

使用分割映射，我们分割训练集合中的每个图像（参看图 2 的方框 14）。然后，我们从每个训练集合图像中的每个分段提取一个或多个特征（参看图 2 的方框 16）。对于术语“特征”，我们表示分段的任何统计量或其他特性。例如，平均像素亮度、中值像素亮度、像素亮度的模式、纹理、直方图、傅立叶变换描述子、小波变换描述子和/或分段中的任何其他统计量。

然后，通过使用特征信息形成分类器（参看图 2 的方框 18）。可使用如现有技术中已知的任何合适类型的分类器。在本发明的具体优

选实施例中，分类器是单类分类器，不需要关于假钞的信息。然而，还可使用二进制分类器或如现有技术中已知的任何合适类型的其他类型的分类器。

图 2 的方法使得用于特定币种和面额的纸币的验证的分类器能够简单、快速、有效地形成。为了创建用于其他币种或面额的分类器，用合适的训练集合图像来重复所述方法。

先前，在 EP1484719 和 US20042447169 中，（如在背景技术部分所提到的）我们使用涉及对图像平面使用网格结构的分割技术和遗传算法方法以形成分割映射。这必然使用关于假钞的一些信息，并且当执行遗传算法搜索时，招致计算成本增加。

这里描述的实施例可使用形成分割映射的不同方法，所述方法不需要使用遗传算法或等效方法，以在大量可能的分割映射中搜索好的分割映射。这减小了计算成本，并提高了性能。另外，不需要关于假钞的信息。

我们认为，在伪造工艺中，通常难以提供整个纸币的一致质量仿效，因此，纸币的特定区域比其他区域更难以成功复制。因此，我们认识到，不使用严格一致的网格分割，我们可通过使用更复杂的分割来提高纸币验证。我们执行指示上述情况的经验检验其实就是这种情况。基于诸如模式、颜色和纹理的形态学特性的分割在检测假钞中导致较好性能。然而，当将诸如使用边缘检测器的传统图像分割方法应用于训练集合中的每个图像时，难以使用所述传统图像分割方法。这是因为对每个训练集合项获得不同的结果，并难以对准不同训练集合图像中的相应特征。为了避免对准分段的问题，在一个优选实施例中，我们使用被称为“时空图像分解”。

现在给出关于形成分割映射的方法的细节。在概要上，该方法可

被认为指定如何将图像平面划分成多个分段，每个分段包括多个指定像素。如上所述，所述分段可以是不连续的。例如，基于来自训练集合中所有图像的信息来编写本说明书。相反，使用严格网格结构的分割不需要来自训练集合中图像的信息。

例如，每个分割映射包括与训练集合中所有图像之间的相应图像元素的关系有关的信息。

认为训练集合中的图像在相同方位相互堆叠并对准。获取纸币图像平面中的给定像素，该像素被认为具有“像素亮度简档（profile）”，所述像素亮度简档包括关于每个训练集合图像中的特定像素位置的像素亮度的信息。使用任何合适的聚类（clustering）算法，将图像平面中的像素位置聚类成分段，在所述分段中的像素位置具有相似或相关的像素亮度简档。

在优选示例中，我们使用这些像素亮度简档。然而，不是必须使用像素亮度简档。还可使用来自训练集合中所有图像的其他信息。例如，4 个相邻像素的块的亮度简档或每个训练集合图像中相同位置的像素的像素亮度的平均值。

现在详细描述我们的形成分割映射的方法的具体优选实施例。这是基于在下面的出版物中教导的方法：Lecture Notes in Computer Science（计算机科学讲义），2352：747-758，2002 中 Avidan, S.的“EigenSegments: A spatio-temporal decomposition of an ensemble of images（特有分段：图像系综的时空分解）”。

给定已经对准并被缩放到相同大小 r×c 的图像系综（ensemble）$\{I_i\}i$=1，2，Λ，N，每个图像 I_i 可通过其像素以向量形式被表示为$[\alpha_{1i}, \alpha_{2i}, \Lambda, \alpha_{Mi}]^T$，其中，$\alpha_{ji}$（j=1，2，Λ，M）是第 i 个图像中第 j 个像素的亮度，M=r·c 是图像中像素的总数量。然后可通过堆叠（stacking）

系综中所有图像的向量 I_i（使用均值归零）产生设计矩阵 $A \in \Re^{M \times N}$，因此，$A=[I_1, I_2, \Lambda, I_N]$。A 中的行向量$[\alpha_{j1}, \alpha_{j2}, \Lambda, \alpha_{jN}]$可被看做 N 个图像的特定像素（第 j 个）的亮度简档。如果两个像素来自图像的相同模式区域，则它们可能具有相似的亮度值，并因此具有强的时间相关性。注意，这里的术语“时间”不需要与时间轴精确对应，但是借用该术语“时间”来指示系综中穿过不同图像的轴。我们的算法尝试找到这些相关性，并将图像平面空间上分割成具有相似时间行为的像素的区域。我们通过定义亮度简档之间的矩阵来测量该相关性。简单方式是使用欧式距离，即两个像素 j 和 k 之间的时间相关性可被表示为 $d(j,k)=\sqrt{\sum_{i=1}^{N}(a_{ji}-a_{ki})^2}$。$d(j,k)$越小，两个像素之间的相关性越强。

为了使用像素之间的时间相关性来在空间上分解图像平面，我们对像素亮度简档（设计矩阵 A 的行）执行聚类算法。这将产生时间上相关像素的簇。最直接的选择是采用 K 均值算法，但可以是任何其他聚类算法。结果，图像平面被分割成时间上相关像素的若干分段。然后，可使用此分段作为模板来分割训练集合中所有图像；并且可关于从训练集合中所有图像的那些分段中提取的特征来构造分类器。

为了不利用假钞而实现训练，优选地是单类分类器。可使用现有技术中已知的任何合适类型的单类分类器。例如，基于神经网络的单类分类器和基于统计的单类分类器。

用于单类分类的合适的统计方法通常是基于在从目标类中提取所考虑的观测值的无效假设下的对数似然比最大化，并且这些方法包括假设目标类为多变量高斯分布的 D^2 检验（在 Morrison，DF: Multivariate Statistical Methods（多变量统计方法）（第三版）中描述。McGraw-Hill 出版公司，纽约，1990）。在任意非高斯分布的情况下，可通过使用例如高斯的半参数混合（在 Bishop，CM: Neural Networks for Pattern Recognition（用于模式识别的神经网络）中描述，牛津大学出版社，纽约，1995）或非参数帕曾（Parzen）窗（在 Duda，RO, Hart， PE, Stork，

DG：Pattern Classification（模式分类）（第二版）中描述，John Wiley 和 Sons, INC, 纽约，2001）来估计目标类的密度，并可通过诸如自抽样（bootstrap）（在 Wang，S，Woodward，WA，Gary，HL 等人：A new test for outlier detection from a multivariate mixture distribution（从多变量混合分布中异常值检测的新检验）中描述, Journal of Computational and Graphical Statistics（计算和图形统计期刊）, 6(3):285-299, 1997）的采样技术来获得无效假设下的对数似然比的分布。

可对单类分类可采用的其他方法是支持向量数据域描述（SVDD）（在 Tax，DMJ，Duin，RPW：Support vector domain description（支持向量域描述）中描述, Pattern Recognition Letters（模式识别快报），20（11-12）：1191-1199，1999）、还有已知的“支持估计（support estimation）”(在 Hayton，P，Scholkopf，B，Tarrassenko，L，Anuzis，P：Support Vector Novelty Detection Applied to Jet Engine Vibration Spectra（应用于喷气式发动机振动谱的支持向量新检测）中描述, Advances in Neural Information Processing Systems（神经网络信息处理系统进阶），13，eds Leen, Todd K 和 Dietterich，Thomas G and Tresp, Volker, MIT Press, 946-952, 2001)和极值理论(EVT) (在 Roberts, SJ: Novelty detection using extreme value statistics（使用极值统计的新检测）中描述. IEE Proceedings on Vision, Image & Signal Processing（关于视觉、图像和信号处理的 IEE 会议记录）, 146(3): 124-129, 1999)。在 SVDD 中，估计支持的数据分布，同时 EVT 估计极值的分布。对此特定应用，用到大量的真钞样本，因此，在这种情况下，可获得目标类分布的可靠估计。因此，在优选实施例中，我们选择可明确估计密度分布的单类分类方法，尽管这不是必要的。在优选实施例中，我们使用基于参数的 D^2 检验的单类分类方法。

在优选实施例中，用于我们的单类分类器的统计量假设检验被详细描述如下：

假定 N 独立，均匀分布的 p 维向量样本（每个纸币的特征集合）

x_1，Λ，$x_N \in C$ 具有关于参数 θ 的基础密度函数 $p(x|\theta)$。对新的点 x_{N+1} 给出下面的假设检验，以使 $H_0: x_{N+1} \in C$，而 $H_1: x_{N+1} \notin C$，其中，C 表示无效假设为真的区域，并通过 $p(x|\theta)$来定义 C。假定备择假设下的分布是均匀的，那么无效和备择假设的正态对数似然比

$$\lambda = \frac{\sup_{\theta \in \Theta} L_0(\theta)}{\sup_{\theta \in \Theta} L_1(\theta)} = \frac{\sup_{\theta} \prod_{n=1}^{N+1} p(\mathbf{x}_n|\theta)}{\sup_{\theta} \prod_{n=1}^{N} (\mathbf{x}_n|\theta)} \qquad (1)$$

可被用作无效假设的检验统计量。在该优选实施例中，我们可使用对数似然比作为用于最新出现的纸币的验证的检验统计量。

1）具有多变量高斯密度的特征向量：假设描述样本中的个别点的特征向量是多变量高斯，从以上的似然比（1）出现的检验评估样本中的每一点是否共享共同均值（在（Morrison，DF: Multivariate Statistical Methods（多变量统计方法）（第三版）中描述。McGraw-Hill 出版公司，纽约，1990））。假定 N 独立，均匀分布的 p 维向量样本 x1，Λ，x_N 来自具有均值 μ 和协方差 C 的多变量正态分布，其样本估计是 $\hat{\mu}_N$ 和 $\hat{C}_N$。样本的随机选择表示为 x_0，相关平方马氏（Mahalanobis）距离

$$D^2 = (\mathbf{x}_0 - \hat{\mu}_N)^T \hat{C}_N^{-1} (\mathbf{x}_0 - \hat{\mu}_N) \qquad (2)$$

可被表示成被分布为具有 p 和 N-p-1 自由度的中心 F 分布

$$F = \frac{(N-p-1)ND^2}{p(N-1)^2 - NpD^2} \qquad (3)$$

然后，如果

$$F > F_{\alpha;p,N-p-1} \qquad (4)$$

则公共总体均值向量 x_0 和剩余 x_i 的无效假设将被拒绝，其中，$F_{\alpha;p,N-P-1}$ 是具有自由度（p，N-p-1）的 F 分布的上 $\alpha \cdot 100\%$ 点。

现在假设选择 x_0 作为具有最大 D^2 统计量的观测值向量。来自大小为 N 的随机样本的最大 D^2 的分布复杂。然而，可通过包法隆尼（Bonferroni）不等式来获得该临界值之上的百分之 100α 的保守近似

值。因此，如果

$$F > F_{\frac{\alpha}{N};p,N-p-1} \tag{5}$$

则我们可断定 x_0 是异常值。

实际上，等式（4）和等式（5）都可用于异常值检测。

当附加数据 x_{N+1} 可用时，在设计没有形成原始采样的部分的新样本的检验中，我们可以使用下面的均值和协方差的增量估计，即均值

$$\hat{\mu}_{N+1} = \frac{1}{N+1}\{N\hat{\mu}_N + \mathbf{x}_{N+1}\} \tag{6}$$

和协方差

$$\hat{C}_{N+1} = \frac{N}{N+1}\hat{C}_N + \frac{N}{(N+1)^2}(\mathbf{x}_{N+1} - \hat{\mu}_N)(\mathbf{x}_{N+1} - \hat{\mu}_N)^T. \tag{7}$$

通过使用表达式（5）和（6）以及矩阵求逆定理，用于 N 个采样参考集合和第 N+1 个检验点的等式（2）变成

$$D^2 = \sigma_{N+1}^T \hat{C}_{N+1}^{-1} \sigma_{N+1} \tag{8}$$

其中，

$$\sigma_{N+1} = (\mathbf{x}_{N+1} - \hat{\mu}_{N+1}) = \frac{N}{N+1}(\mathbf{x}_{N+1} - \hat{\mu}_N) \tag{9}$$

$$\hat{C}_{N+1}^{-1} = \frac{N+1}{N}\left(\hat{C}_N^{-1} - \frac{\hat{C}_N^{-1}(\mathbf{x}_{N+1} - \hat{\mu}_N)(\mathbf{x}_{N+1} - \hat{\mu}_N)^T \hat{C}_N^{-1}}{N+1+(\mathbf{x}_{N+1} - \hat{\mu}_N)^T \hat{C}_N^{-1}(\mathbf{x}_{N+1} - \hat{\mu}_N)}\right) \tag{10}$$

通过 $D_{N+1,N}^2$ 来表示 $(\mathbf{x}_{N+1} - \hat{\mu}_N)^T \hat{C}_N^{-1}(\mathbf{x}_{N+1} - \hat{\mu}_N)$，则

$$D^2 = \frac{ND_{N+1,N}^2}{N+1+D_{N+1,N}^2} \tag{11}$$

所以，可依据普通估计的均值 $\hat{\mu}_N$ 和协方差 $\hat{C}_N$ 的估计和假设的正态分布来检验新点 x_{N+1}。尽管发现多变量高斯特征向量的假设对许多应用是合适的实用选择，但多变量高斯特征向量的假设实际上通常不成立。在下面的部分中，我们放弃这种假设，并考虑任意密度。

2）具有任意密度的特征向量：通过使用现有技术中已知的任何合适的半参数（例如，高斯混合模型）或非参数（例如，帕曾窗方法）密度估计方法，可从自任意密度 $p(\mathrm{x})$中提取的有限数据样本 $S=\{\mathrm{x}_1,\Lambda,\mathrm{x}_N\}\in\Re^d$ 获得概率密度估计 $\hat{p}(x;\theta)$。然后，在计算对数似然比（1）中可采用该密度。与多变量高斯分布的情况不同，在无效假设下，不存在检验统计量（λ）的分析分布。所以，为了获得这种分布，可采用数字自抽取方法来获得估计的密度下的另外的非分析无效分布，所以，可从获得的经验分布中建立各种临界值 λ_{crit}。可以看出，在极限值 $N\to\infty$，似然比可通过以下来估计

$$\lambda=\frac{\sup_{\theta\in\Theta}L_0(\theta)}{\sup_{\theta\in\Theta}L_1(\theta)}\to\hat{p}\left(\mathbf{x}_{N+1};\hat{\theta}_N\right) \tag{12}$$

其中，$\hat{p}(x_{N+1};\hat{\theta}_N)$表示在通过原始 N 个样本估计的模型下的 x_{N+1} 的概率密度。

在从参考数据集合产生 N 个样本的 B 集合自抽样并使用其估计密度分布的参数 $\hat{\theta}_N^i$ 之后，可通过随机选择第 N+1 个样本并计算 $\hat{p}(x_{N+1};\hat{\theta}_N^i)\approx\lambda_{crit}^i$ 来获得 B 自抽样复制的检验统计量 $\lambda_{crit}^i, i=1,K,B$。通过以升序的方式对 λ_{crit}^i 排序，可限定临界值 α，从而如果 $\lambda\le\lambda_\alpha$，则在期望的有效等级拒绝无效假设，其中，λ_α 是 λ_{crit}^i 的第 j 个最小值，并且 $\alpha=j/(B+1)$。

优选地，为不同的分段数，重复形成分类器的方法，并使用已知真伪的纸币的图像进行检验。然后，选择给出最佳性能的分段的数量，并且使用所述数量的分段的分类器被使用。我们发现，尽管可使用任何合适数量的分段，但是分段的最佳数量大约是从 2 到 15。

如上所述，特定问题涉及识别和代替将被验证的纸币的图像中的异常图像元素。图 3 是用决策-中性数据代替异常图像元素的处理的流

程图。对于每个图像元素（方框 300），例如像素、像素组，获取图像定位的分布（方框 301）。所述分布是对所述图像定位遍及在图像的训练集合的所有图像的估计分布。如上所述，图像的训练集合可以是真钞的多个图像。例如，所述分布可以是像素亮度简档或四个像素定位的块的亮度简档，或如上述类似。优选地，所述分布与在如上所述形成用于纸币验证器的分割映射的处理期间所使用的分布相同。因为已经估计了所述分布，所以这减小了计算成本，并节省了时间。

然后，基于置信度（还被称为置信度）从获取的分布中选择值（方框 302）。所述置信度与在纸币验证器中使用的分类器的置信度有关。例如，所述置信度与分类器使用的置信度相同。因为置信度与分类器的置信度有关，所以通过以这种方式选择值来获得决策-中性数据。然后，用所选择的值来代替异常图像元素处的值（参看方框 303）。通过以这种方式使用决策-中性数据，我们保证纸币的剩余部分指示纸币验证器的分类结果。这优于传统方法，对于所述传统方法，真钞上缺失或不可靠的数据意味着，为避免许多错误拒绝而将遭受错误接受率。以这种方式，我们无需修改核心纸币验证处理就能够成功地解决损坏的、破旧的、有破缝的或部分退色的纸币。所需要纸币图像的只是预处理。另外，这在不对错误接受率折衷的情况下被实现。

图 4 是创建用于纸币验证的分类器 22 的设备 20 的示意图。它包括：

- 输入 21，配置以访问纸币图像的训练集合；
- 处理器 23，配置以使用训练集合图像创建分割映射；
- 分割器 24，配置以使用分割映射分割每个训练集合图像；
- 特征提取器 25，配置以从每个训练集合图像的每个分段中提取一个或多个特征；以及
- 分类形成装置 26，配置以使用特征信息形成分类器；

其中，配置处理器以基于来自训练集合中所有图像的信息来创建分割映射。例如，通过使用上面描述的空时图像分解。

图 5 是纸币验证器 31 的示意图。它包括：

- 输入，配置以接收将被验证的纸币的至少一个图像 30；
- 分割映射 32；
- 处理器 36，配置以识别图像中的异常；
- 图像修改器 37，配置以通过用中性决策生成数据代替识别出的异常来形成修改的图像，作为中性决策生成数据的所述数据与分类器 35 有关；
- 另一处理器 33（可与处理器 36 集成），配置以使用分割映射分割纸币的图像；
- 特征提取器 34，配置以从纸币图像的每个分段中提取一个或多个特征；
- 分类器 35，配置以基于提取的特征将纸币分类为有效或无效；

其中，分割映射包括与纸币的图像的训练集合中所有图像之间的相应图像元素的关系有关的信息。注意，对于图 5 的装置不必相互独立，这些装置可以是集成的。

图 6 是验证纸币的方法的流程图。所述方法包括：

- 访问将被验证的纸币的至少一个图像（方框 40）；
- 识别异常图像元素（方框 41）；
- 用决策中性数据代替异常图像元素（方框 42）；
- 访问分割映射（方框 43）；
- 使用分割映射分割纸币的图像（方框 44）；
- 从纸币图像的每个分段中提取特征（方框 45）；
- 使用分类器基于所提取的特征将纸币分类为有效或无效（方框 46）；

其中，分割映射是基于与纸币的训练图像集合中的每个图像有关的信息形成的。可以以现有技术中已知的任何合适的顺序或组合执行这些方法的步骤。分割映射可暗含包括与训练集合中每个图像有关的信息，这是因为可基于所述信息形成分割映射。然而，在分割映射中暗含的

信息可以是具有像素地址列表的简单文件，其将被包括在每个分段中。

图 7 是具有纸币验证器 53 的自助式设备 51 的示意图。它包括：

- 用于接受纸币的装置 50；
- 用于获得纸币的数字图像的成像装置 52；
- 用于用决策-中性数据代替异常图像的处理器 54；以及
- 如上所述的纸币验证器 53。

可对纸币的图像或其他表现执行这里描述的方法，所述图像/表现是任何合适类型。例如，红色、蓝色和绿色通道上的图像或如上所述的其他图像。

可只基于一种类型（比如红色通道）的图像来形成分割。可选择地，可基于所有类型（比如红色、蓝色和绿色通道）的图像来形成分割映射。还可形成多个分割映射，对每种图像或多种图像类型的组合形成一种分割映射。例如，可以有三个分割映射，一个分割映射用于红色通道图像，一个分割映射用于蓝色通道图像，一个分割映射用于绿色通道图像。在这种情况下，在单个纸币的验证期间，依据所选图像的类型使用合适的分割映射/分类器。因此，可通过使用不同类型的图像和相应的分割映射/分类器来修改上述方法中的每一种方法。

与成像装置一样，用于接受纸币的装置是现有技术中已知的任何合适类型。可使用现有技术中已知的任何特征选择算法来选择一种或多种特征以在提取特征步骤中使用。另外，除这里讨论的特征信息之外，还可基于与纸币的特定面额或币种有关的特定信息来形成分类器。例如，与在颜色方面数据显著丰富的区域的信息、在给定币种和面额中的形状或空间频率。

对于技术人员来说是显然的，在不丧失效果的情况下，可扩展或改变这里所给出的任何范围或装置值。

应该理解，只作为示例给出了优选实施例的以上描述，本领域的技术人员可进行各种修改。

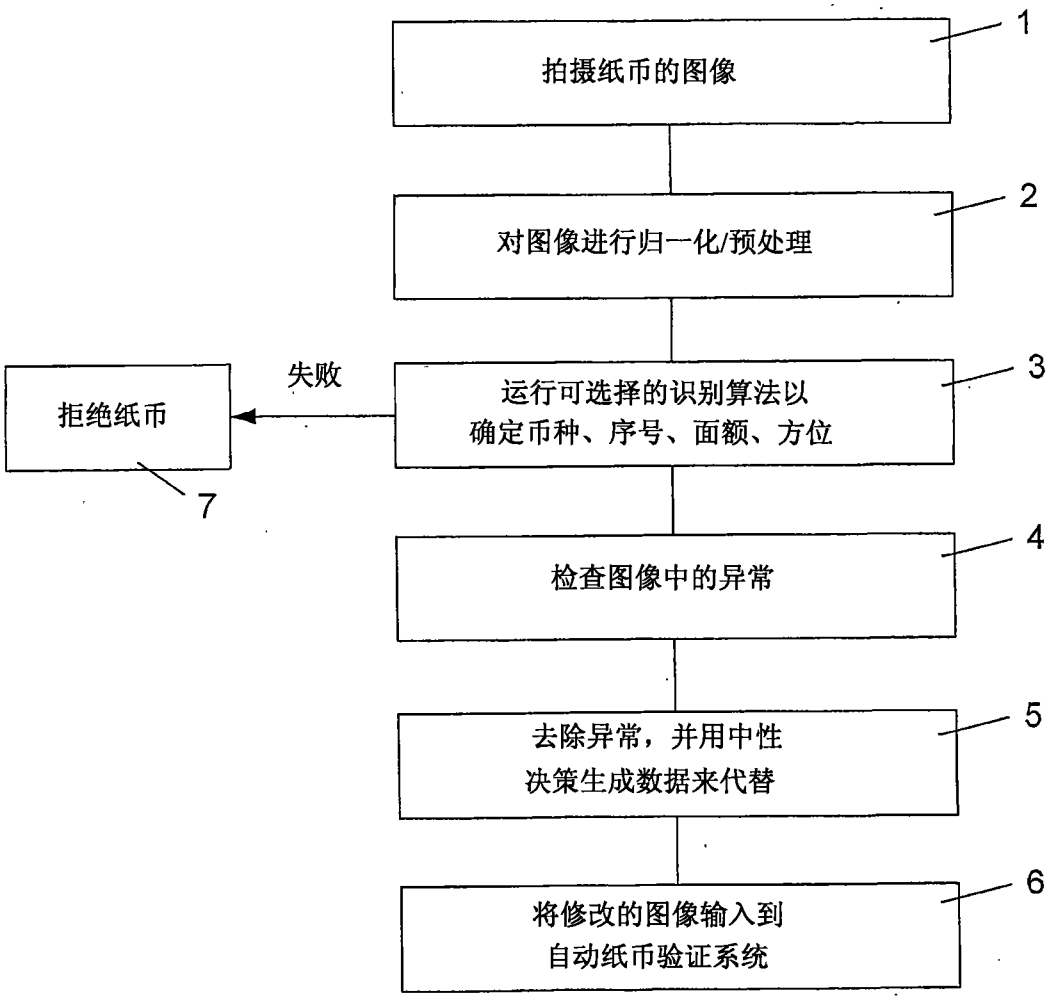

拍摄纸币的图像
1
对图像进行归一化/预处理
2
运行可选择的识别算法以
确定币种、序号、面额、方位
3
失败
拒绝纸币
7
检查图像中的异常
4
去除异常，并用中性
决策生成数据来代替
5
将修改的图像输入到
自动纸币验证系统
6

图1

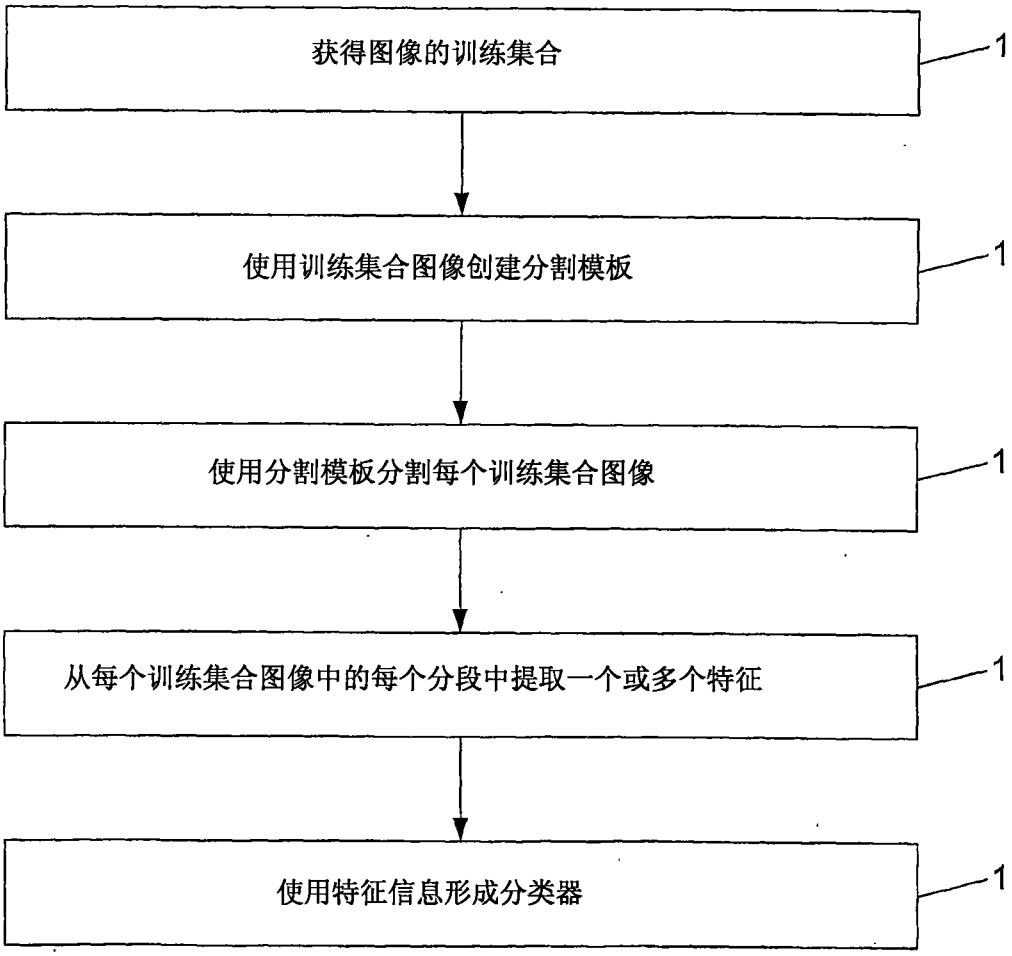

获得图像的训练集合
1
使用训练集合图像创建分割模板
1
使用分割模板分割每个训练集合图像
1
从每个训练集合图像中的每个分段中提取一个或多个特征
1
使用特征信息形成分类器
1

图2

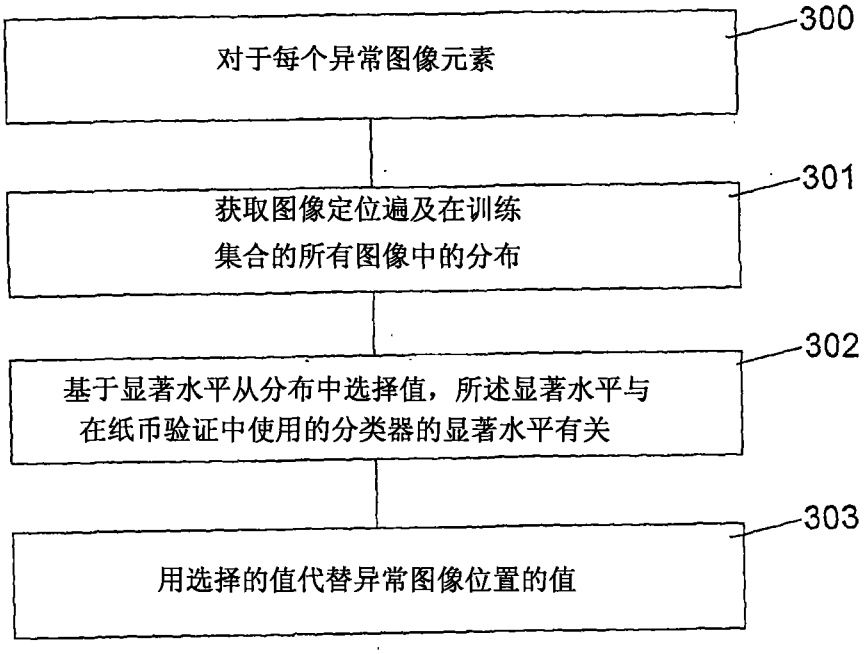

对于每个异常图像元素
300
获取图像定位遍及在训练
集合的所有图像中的分布
301
基于显著水平从分布中选择值，所述显著水平与
在纸币验证中使用的分类器的显著水平有关
302
用选择的值代替异常图像位置的值
303

图3

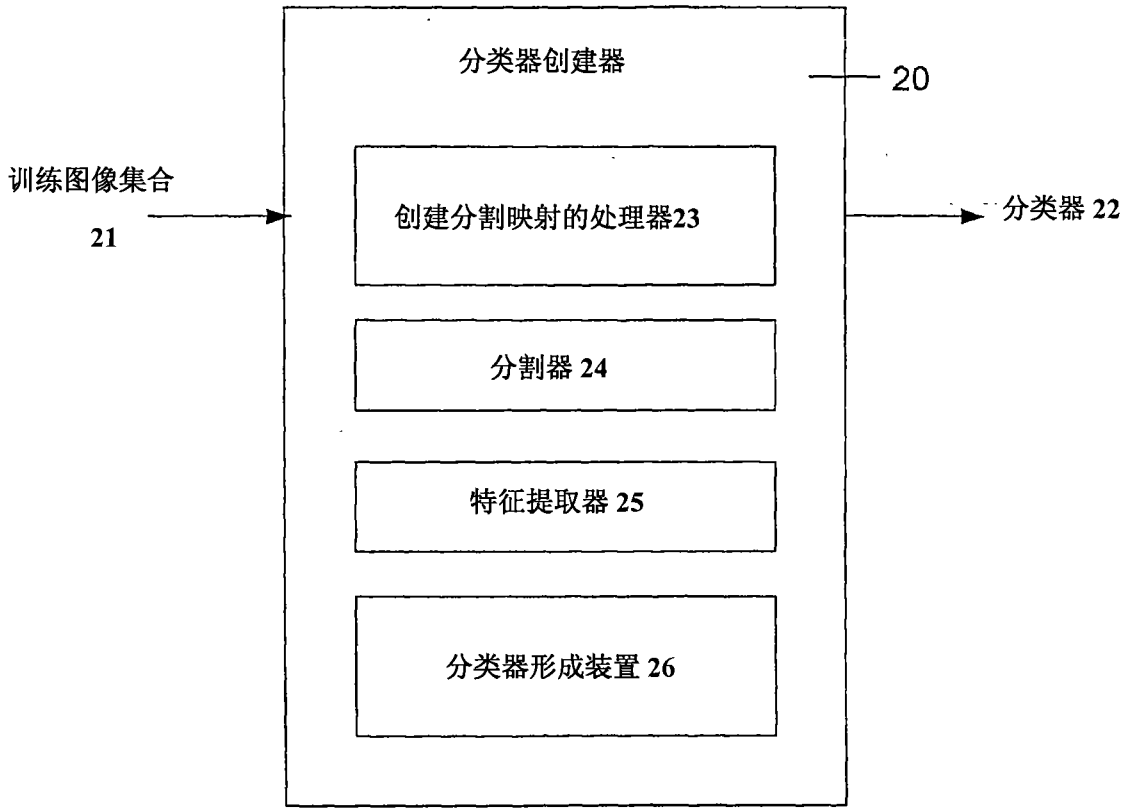

分类器创建器
20
训练图像集合
21
创建分割映射的处理器23
分类器 22
分割器 24
特征提取器 25
分类器形成装置 26

图4

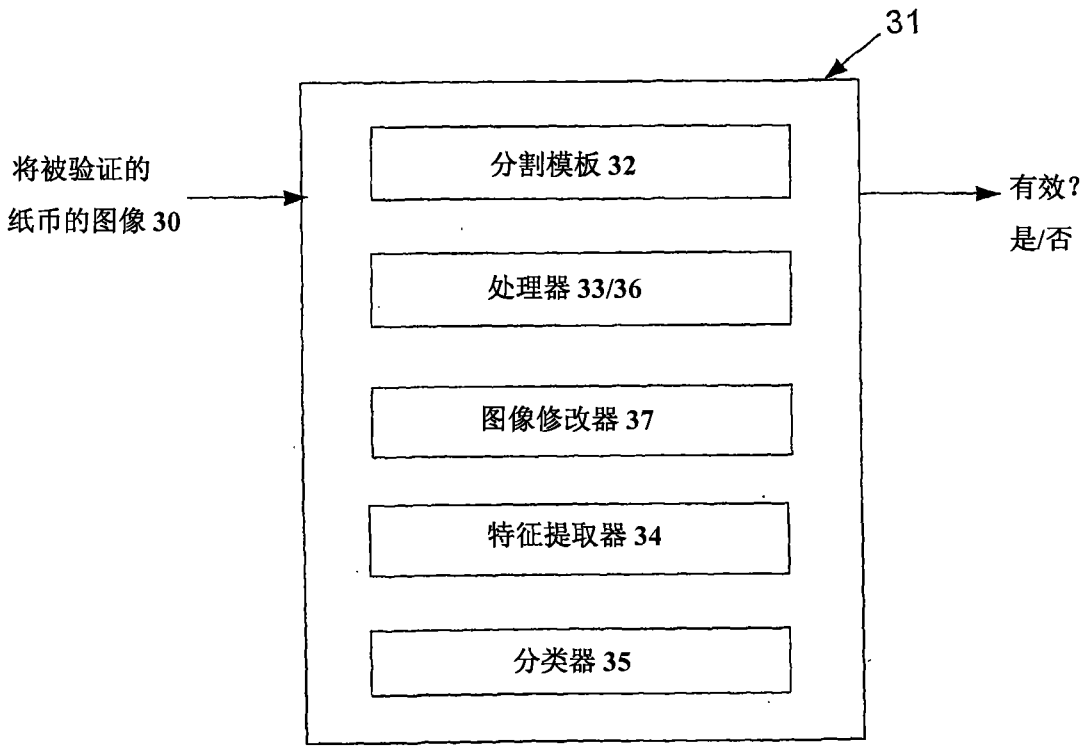

31
将被验证的
纸币的图像 30
分割模板 32
处理器 33/36
图像修改器 37
特征提取器 34
分类器 35
有效?
是/否

图5

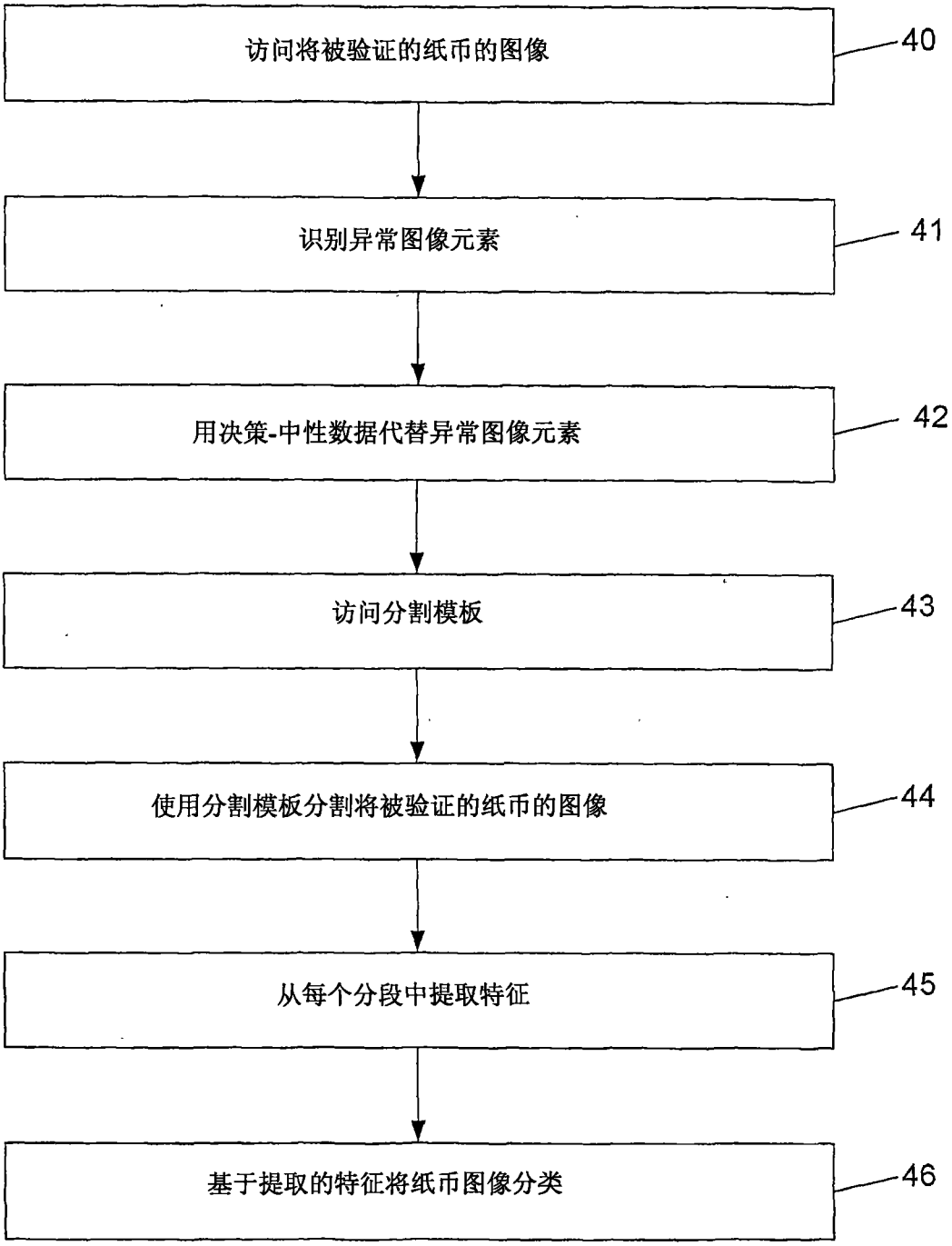

访问将被验证的纸币的图像
40
识别异常图像元素
41
用决策-中性数据代替异常图像元素
42
访问分割模板
43
使用分割模板分割将被验证的纸币的图像
44
从每个分段中提取特征
45
基于提取的特征将纸币图像分类
46

图6

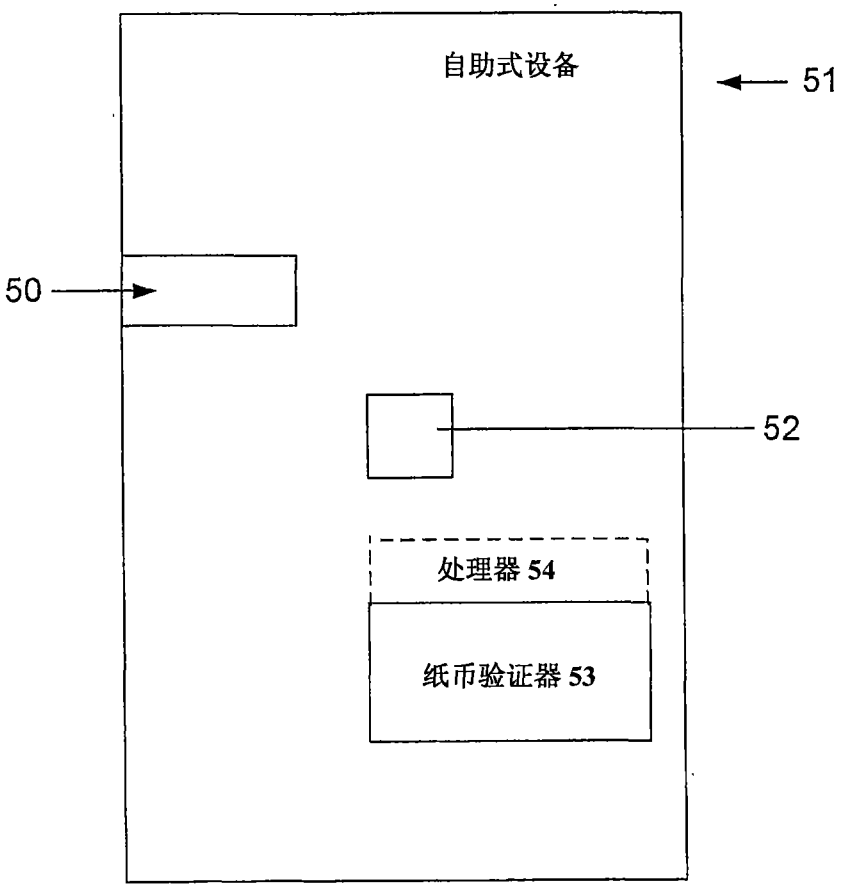


图7