

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

H04N 1/32 (2006.01)

G06T 1/00 (2006.01)



[12] 发明专利申请公布说明书

[21] 申请号 200910127164.X

[43] 公开日 2009 年 11 月 25 日

[11] 公开号 CN 101588433A

[22] 申请日 2005.1.6

[21] 申请号 200910127164.X

分案原申请号 200580002034.3

[30] 优先权

[32] 2004.1.6 [33] US [31] 60/534,520

[71] 申请人 汤姆逊许可证公司

地址 法国布洛尼-比扬古市

[72] 发明人 贾斯廷·皮卡德 赵 健

[74] 专利代理机构 北京东方亿思知识产权代理有限公司

代理人 宋 鹤 南 霆

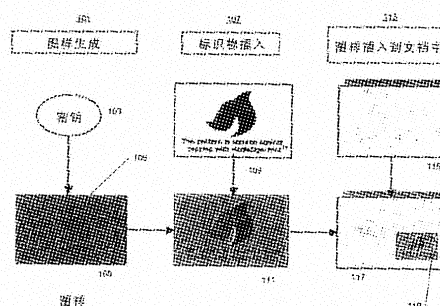
权利要求书 3 页 说明书 50 页 附图 18 页

[54] 发明名称

用于检测、分析和使用可见认证图样的改进技术

[57] 摘要

公开了用于检测、分析和使用可见认证图样的改进技术。用于制作拷贝检测图样并用它们来检测数字表示形式之间的拷贝关系的改进技术。这些技术包括：在不更改拷贝检测图样平均信息量的情况下将消息包括在拷贝检测图样中和读取消息，在不参考模拟形式的拷贝检测图样的原本的数字表示形式的情况下利用模拟形式上的拷贝检测图样确定模拟形式是否原始模拟形式，通过修改比较的两个拷贝检测图样之一以考虑由拷贝过程导致的更改来增大利用拷贝检测图样进行拷贝检测的灵敏性，将拷贝检测图样分布在文档上，以及利用拷贝检测图样的平均信息量来定位拷贝检测图样。还公开了涉及数字形式和模拟形式之间的变换的拷贝和数字-数字拷贝对拷贝检测图样的应用。



1. 一种数字表示形式，该数字表示形式包括内容，并且其特征在于：

所述数字表示形式还包括拷贝关系部分，该拷贝关系部分包含由产生了所述数字表示形式的拷贝过程造成的假象，并且所述拷贝关系部分被用于确定所述数字表示形式和包括所述内容的另一数字表示形式之间的拷贝关系。

2. 如权利要求 1 所述的数字表示形式，其特征还在于：

所述数字表示形式的所述拷贝关系部分被用于确定所述数字表示形式相对于所述另一数字表示形式的拷贝世代。

3. 如权利要求 2 所述的数字表示形式，其特征还在于：

所确定的拷贝世代是第一拷贝世代。

4. 如权利要求 1 所述的数字表示形式，其特征还在于：

所述数字表示形式的所述拷贝关系部分被用于确定所述数字表示形式是否是经授权的。

5. 如权利要求 1 所述的数字表示形式，其特征还在于：

所述数字表示形式的所述拷贝关系部分被用于获得关于所述拷贝过程的信息。

6. 如权利要求 1 至 5 中任何一项所述的数字表示形式，其特征还在于：

所述拷贝过程是数字的；并且

在所述拷贝过程期间所述内容经历数字差错校正，而所述拷贝关系部分没有经历数字差错校正。

7. 如权利要求 1 至 5 中任何一项所述的数字表示形式，其特征还在于：

所述拷贝过程包括从所述另一数字表示形式产生模拟输出并且从所述模拟输出产生所述数字表示形式。

8. 一种确定包括内容的数字表示形式和包括该内容的另一数字表示形式之间的拷贝关系的方法，所述另一数字表示形式还包括拷贝关系部分，

该拷贝关系部分包含由产生了所述另一数字表示形式的拷贝过程造成的假象，并且所述拷贝关系部分被用于确定所述数字表示形式和所述另一数字表示形式之间的拷贝关系，并且所述方法的特征在于以下动作：

接收所述另一数字表示形式；

利用所述另一数字表示形式的拷贝关系部分来确定所述拷贝关系。

9. 如权利要求 8 所述的方法，其特征还在于：

所确定的拷贝关系是所述另一数字表示形式相对于所述数字表示形式的拷贝世代。

10. 如权利要求 8 所述的方法，其特征还在于：

所述拷贝世代是第一拷贝世代。

11. 如权利要求 8 所述的方法，其特征还在于：

所确定的拷贝关系指示出所述另一数字表示形式是否是经授权的。

12. 如权利要求 8 所述的方法，其特征还在于：

所述拷贝关系指示出所述拷贝过程。

13. 如权利要求 8 所述的方法，其特征还在于：

所述方法的动作是在对所述另一数字表示形式执行操作的设备中执行的；并且

所述方法还包括以下动作：

如果所确定的拷贝关系指示出在所述另一数字表示形式上许可所述操作，则对所述另一数字表示形式执行所述操作。

14. 如权利要求 13 所述的方法，其特征还在于：

所述操作是从所述另一数字表示形式产生模拟输出。

15. 如权利要求 13 所述的方法，其特征还在于：

所述操作是制作另外的数字表示形式，该另外的数字表示形式是所述另一数字表示形式的拷贝。

16. 如权利要求 8 至 15 中任何一项所述的方法，其特征还在于：

所述拷贝过程是数字的；并且

在所述拷贝过程期间所述内容经历数字差错校正，而所述拷贝关系部分没有经历数字差错校正。

17. 如权利要求 8 至 15 中任何一项所述的方法，其特征还在于：

所述拷贝过程包括从所述数字表示形式产生模拟输出并且从所述模拟输出产生所述另一数字表示形式。

18. 一种用于从数字表示形式产生模拟输出的装置，该数字表示形式包括拷贝关系部分，该拷贝关系部分包括由产生了所述数字表示形式的拷贝过程造成的假象，并且所述装置的特征在于：

拷贝关系检测器，该拷贝关系检测器根据所述拷贝关系部分确定所述数字表示形式是否是经授权的；以及

模拟输出产生器，如果所述拷贝关系检测器确定了所述数字表示形式是经授权的，则该模拟输出产生器从所述数字表示形式产生所述模拟输出。

19. 一种用于拷贝数字表示形式的装置，该数字表示形式包括拷贝关系部分，该拷贝关系部分包括由产生了所述数字表示形式的拷贝过程造成的假象，并且所述装置的特征在于：

拷贝关系检测器，该拷贝关系检测器根据所述拷贝关系部分确定所述数字表示形式是否是原始数字表示形式；以及

拷贝器，如果所述拷贝关系检测器确定了所述数字表示形式是原始数字表示形式，则该拷贝器产生所述数字表示形式的数字拷贝。

用于检测、分析和使用可见认证图样的改进技术

本申请是 2005 年 1 月 6 日递交的发明名称为“用于检测、分析和使用可见认证图样的改进技术”、申请号为 200580002034.3 (PCT/US2005/000201) 的专利申请的分案申请。

与相关申请的交叉引用

本申请要求 2004 年 1 月 6 日递交的与本发明有相同发明人的标题为“*Visible Authentication Patterns for Printed Documents – Extensions II*”的美国临时申请 60/534,520 的优先权。本申请还通过全文引用结合了 2004 年 11 月 12 日递交的 Picard, Zhao 和 Thorwirth 的标题为“*Visible Authentication Patterns for Printed Documents*”的 U.S.S.N. 10/514,271, 用于所有目的。U.S.S.N. 10/514,271 是 2003 年 11 月 27 日作为 WO 03/098540 公布的 PCT/US03/15168 的美国国家阶段。本专利申请的“具体实施方式”包含 U.S.S.N. 10/514271 的“具体实施方式”的实质内容。本申请的“具体实施方式”的新材料从标题为“检测 VAP 的位置”的部分开始。

技术领域

本发明一般地涉及打印文档中的安全性特征, 更具体而言, 涉及打印文档中的可见认证图样。可见认证图样可被用于区分原始打印文档和这些打印的文档的影印拷贝, 以检测文档的更改并且携带隐藏的和/或可见的消息。

背景技术

可见认证图样 (Visible Authentication Pattern, VAP) 可被用于确定打印的文档是否已经被更改或文档是原本 (original) 还是拷贝 (copy)。VAP 是数字文档的一部分中的有噪声 (noisy) 图样。它被用于通过将模

拟形式作出的数字记录的一部分与模拟形式的该部分的原始数字表示形式相比较以确定记录的部分和该部分的原始数字表示形式之间的相异（或相似）程度并且利用所述相异（或相似）程度确定模拟形式是否是原始模拟形式，从而来确定文档的真实性。

VAP 及其使用是上述美国专利申请 10/514,271 的主题。有关 VAP 的更进一步经验引起了多个领域的改进：

- 在不影响 VAP 区分拷贝和原本的属性的情况下在 VAP 中存储信息；
- 利用 VAP 的平均信息量（entropy）在文档图像中定位它；
- 减轻将 VAP 结合到文档中的审美影响；
- 应对以下事实：VAP 可能经历多种打印和扫描过程，并且这些过程会修改出现在原始模拟拷贝和非原始模拟形式上的 VAP；
- 在涉及数模变换和模数变换甚至数字到数字拷贝的其他应用中使用 VAP 的类似物。

这里所公开的发明的一个目的是提供就这些和其他领域来说有所改进的 VAP。

发明内容

在一个方面中，本发明的目的是通过用于在对数字认证图样的平均信息量产生最小限度的影响的情况下将消息结合到数字认证图样中的技术来实现的。在该技术中，VAP 中的多组图样元素携带消息的消息元素。在携带消息元素的每组图样元素中，图样元素的值被设置，以便它们代表消息元素。设计图样元素的方式对于图样元素的平均信息量和它们的其他所需属性具有最小限度的影响。本发明的这个方面包括用于创建具有消息的数字认证图样和用于读取消息的方法和装置，以及利用该技术制成的包含消息的数字认证图样。

在另一个方面中，本发明的目的是通过用于在不参考模拟形式的拷贝检测图样的原本的数字表示形式的情况下确定包括拷贝检测图样的模拟形式是否是原始模拟形式的技术来实现的。该技术扫描模拟文档上的拷贝检

测图样，以制作拷贝检测图样的数字表示形式，并利用数字表示形式的一个或多个全局属性来制作初步确定模拟形式是否是原始模拟形式。

另一个方面是用于确定数字表示形式之间的拷贝关系的技术。数字表示形式中的每一个包括对由拷贝过程产生的更改敏感的部分，该技术修改来自数字表示形式之一的该部分，以使它更可以与来自其他数字表示形式的该部分相比拟。修改考虑到了由拷贝过程产生的更改。当进行修改时，这些部分被比较，以确定数字表示形式之间的拷贝关系。

其他方面包括：

- 用于文档的可视认证图样，其被细分为分布在文档上的多个单元；
- 用于利用可视认证图样的平均信息量来定位图样的技术；
- 模拟信号的数字表示形式，其包括对由模数和数模转换产生的变换敏感的拷贝检测信号的一种表示形式；以及
- 包括第一部分和第二部分的数字表示形式，在第一部分中数据具有差错校正，在第二部分中数据没有差错校正并且数据对于由制作数字表示形式的数字拷贝的过程产生的更改敏感。

本发明相关领域的技术人员在熟读下面的“具体实施方式”和附图之后，将会清楚看到本发明的其他目的和优点。

附图说明

图 1 是关于如何生成可见认证图样（VAP）并将其插入到文档中的概况；

图 2 示出如何从文档记录 VAP；

图 3 是示出如何将 VAP 用于认证中的概况的流程图；

图 4 是原始和非原始模拟形式的打印和认证的概况；

图 5 示出用于水印检测和更改检测的 GUI；

图 6 是示出 VAP 的原始数字表示形式和从非原始文档记录的 VAP 的频带中的能量之间的相互关系的图；

图 7 是示出 VAP 的原始数字表示形式和从原始文档记录的 VAP 的频带中的能量之间的相互关系的图；

图 8 示出如何用基于消息的密钥（key）来将无内容水印嵌入到图像中；

图 9 示出如何将 VAP 结合到条形码或标识物（logo）中；功能一般是固定的，并且在每个测试 CDP 被恢复之后被应用到其上；

图 10 示出示例性直方图转换函数；

图 11 示出被扫描的经恢复的 CDP 上的直方图转换的作用；

图 12 示出利用不同密钥生成的若干个 CDP；

图 13 示出具有平均分布的模板直方图；

图 14 是利用分布式 CDP 进行的检查；

图 15 是 CDP 的校准图像；

图 16 示出分布在整个文档中或文档内的对象中的 CDP；

图 17 示出插入了消息的一组像素的变换；以及

图 18 示出如何将 VAP 用于可视加密。

图中的标号有三个或更多个数字：右边两个数字是由剩余的数位所指示的图之中的标号。从而，标号为 203 的项目首先作为项目 203 出现在图 2 中。

具体实施方式

以下直到“VAP 与其他安全性技术的结合”部分为止的“具体实施方式”是从上述 Picard, Zhao 和 Thorwirth 的标题为“*Visible Authentication Patterns for Printed Documents*”的 U.S.S.N. 10/514,271 获得的。正如该专利申请中所说明的，数字水印在文档中的一种用途是用来检测由于打印-扫描过程导致的修改，并从而确定文档是原本还是拷贝。另一个用途是用来检测文档是否已被更改。水印的这两个用途的特性是它们是基于水印的存在与否的，而不是基于其内容的。

可见认证图样

当仅仅是水印的存在性被用于确定模拟形式的真实性时，水印被用作无内容图样。由于图样没有内容，因此它就不需要是不可见的；相反，它

可以作为可见元素被添加到文档中。在下文中，用于认证的可见图样被称为*可见认证图样*或 VAP。由于 VAP 是可见的，因此比起水印来它容易检测得多，并且由于原始文档的拷贝而导致的恶化量可以被更好地估计。但是，它仍能够执行不可见水印的所有认证功能，并且还能使文档的消费者知道文档的真实性是受到保护的。

术语

下面的术语将被用于“*具体实施方式*”中，以阐明数字表示形式和模拟形式之间的关系。

对象的*数字表示形式* (*digital representation*) 是对象的一种形式，在这种形式中，对象可以被存储在数字处理系统中，并且可以被数字处理系统操纵。对象可以是或者可以包括文档、图像、音频、视频或其数字表示形式能够被制成的任何其他媒体。

数字表示形式的*模拟形式* (*analog form*) 是在数字表示形式被输出到模拟设备时产生的对象或成分的形式，所述模拟设备例如是显示器、打印机、扬声器、烧制器（高密盘或 DVD 盘）、雕刻 (*engraving*) 或压纹 (*embossing*) 设备。

模拟形式的*数字记录* (*digital recording*) 是从模拟形式制成的数字表示形式。制作数字记录的方式依赖于媒体；例如，对于文档或图像，数字记录是通过对从文档或图像的模拟形式制成的图像进行数字化来完成的。

原始数字表示形式 (*original digital representation*) 是由被授权者制成或重新创建成的数字表示形式；*原始模拟形式* (*original analog form*) 是从原始数字表示形式制成的模拟形式。

非原始数字表示形式 (*non-original digital representation*) 是通过数字地记录模拟形式而制成的数字表示形式；*非原始模拟形式* (*non-original analog form*) 是从非原始数字表示形式制成的或通过对模拟形式进行影印而制成的模拟形式。

文档 (*document*) 将会被赋予以下特殊意义：由打印过程产生的任何模拟形式，其中包括通常意义上的自身带有特征的文字、标志、包装和对

象。这里所使用的打印（*printing*）包括例如雕刻或压纹这样的过程。在可以进行合理类推的范围内，下面所说的任何关于文档的内容都同样适用于其他媒体。例如，音频模拟形式可以包括作为 VAP 的音频等同物的可听认证图样。

制作可见认证图样：图 1

可见认证图样的矛盾之处在于虽然图样是可见的，但可能的造假者必须不能修改图样以使它认证不真实的文档。在优选实施例中，这一目标是通过使图样有噪声来实现的，即构成图样的像素的值的大部分明显是随机确定的。由于图样是有噪声的，因此如果不访问图样的原始数字表示形式就不可能辨别构成图样的数字表示形式的像素应当具有什么样的值。另一方面，在给定 VAP 的原始数字表示形式的情况下，可以将来自文档的 VAP 的数字记录与 VAP 的原始数字表示形式相比较，确定记录的 VAP 相对于 VAP 的原始数字表示形式经历了怎样的更改，并且可以从差异确定所讨论的文档经历了怎样的更改。从下面可以更详细看出，能够检测到的更改包括在制作非原始文档时涉及的更改以及在更改文档中的信息时涉及的更改。

图 1 示出制作可见认证图样并将其插入到文档中的一种方式。有三个步骤：

- 生成图样的数字表示形式，如 101 处所示；
- 可选的步骤：将可见标识物或图例添加到认证图样中，如 107 处所示；以及
- 将认证图样插入到文档中，如 113 处所示。

图样 105 的原始数字表示形式可以用任何产生以下结果的方式来生成：图样的像素看起来拥有具有强随机成分的值。图样 105 的数字表示形式可以是灰度级图样，或者它可以采用彩色像素。采用密钥来生成图样尤其有用；密钥 103 被用作伪随机数生成器的种子，该生成器产生被赋给图样中的像素的值的序列。后面将详细说明密钥的使用。图样 105 的原始数字表示形式还可以包括帮助在通过扫描包含图样 105 的文档制成的数字表

示形式中定位图样的成分。在图样 105 中，黑边 106 执行该功能。

可见标识物或图例 109 可以被添加到图样 105 的原始数字表示形式，以便在不损害图样 105 的噪声性的情况下制成图样 111 的原始数字表示形式，这是因为构成图样的像素的值中只有一部分需要随机的确定。从而，可以通过以保持标识物或图例的随机性同时使得标识物或图例显现的方式操纵构成标识物或图例的像素的值来将标识物或图例叠加在图样 105 中。例如，如果图样 105 是灰度级图样，则可以通过使图例或标识物的像素相对于其原始随机值均匀地更暗或更亮来制作图例或标识物。该技术与向图像添加可见水印类似，只不过它保持了图样 105 的噪声性。

一旦图样 111 的原始数字表示形式已经被制成，它就被插入到文档 115 的原始数字表示形式中，如 113 处所示。当从原始数字表示形式 115 打印文档 117 时，文档 117 包括打印的可见认证图样 119。当然，文档可以被打印到其上已经有打印的材料的基础上。从而，图样 119 可以被添加到预打印的基底。

利用可见认证图样来认证文档；图 2 和图 3

当包含打印的 VAP 119 的文档被认证时，发生下面步骤：

- 在文档中检测到打印的 VAP 119.
- 制作检测到的打印的 VAP 119 的数字记录。
- 将打印的 VAP 的数字记录与 VAP 的原始数字表示形式相比较；以及
- 基于比较确定真实性。

打印的 VAP 的数字记录与 VAP 的原始数字表示形式相比较的方式依赖于所进行的认证的种类；此外，对给定文档的认证可以包括在数字记录和原始数字表示形式之间进行的若干个不同种类的比较。例如，支票的数量字段上的可见认证图样的数字记录可以首先与原始数字表示形式相比较，以确定支票是否是伪造品，然后确定数量字段中的数量是否被更改过。

图 2 示出优选实施例中的检测打印的 VAP 和制作 VAP 的数字记录。

这两者都是利用可从 MediaSec Technologies 获得的“Scanread”应用程序完成的。也可以采用检测文档的一部分并制作其数字记录的其他应用。Scanread 201 利用黑边 106 来检测打印的文档 117 中可见认证图样 119 的存在，然后制作可见认证图样 119 的数字记录 203。图 3 示出了利用数字记录 203 和 VAP 119 的原始数字表示形式 111 来确定真实性的程序的一般流程图 301。VAP 的原始数字表示形式 111 可以是原本本身、原本的拷贝，或者是用与第一原始数字表示形式完全相同的方式制作的新的原始数字表示形式 111。通过这些方法中的任何一种获得的原始数字表示形式当然是完全等同的，使用哪种方法是实现方式方面的问题，例如 VAP 的原始数字表示形式的存储成本、在网络上传输 VAP 的原始数字表示形式的成本，以及每次需要时生成原始数字表示形式的成本。

从 303 处开始，在 305 处，数字记录 203 和原始数字表示形式 111 的特征被比较；比较哪些特征以及如何比较这些特征依赖于所进行的认证的种类。如果数字记录 203 和原始数字表示形式 111 之间的差异超过阈值（307），则存在认证问题，并且取分支 309。阈值也将依赖于所进行的认证的种类。在分支 309 中，在 311 处，向进行认证的应用程序指示问题的存在。如果有用的话，程序还可以提供关于比较的信息（315）；同样，信息的各种和提供信息的方式将会依赖于认证种类。例如，如果数量字段中的数量看起来曾被更改，则程序可以显示一幅图像，该图像示出在可见认证图样的数字记录中原始数字表示形式的哪些像素看起来曾被更改。如果差异不超过阈值，则取分支 317。在这里，向进行认证的应用程序指示未检测到认证问题这一事实。两条分支和程序都终止于 321 处。

利用可见认证图样来区分原始文档和非原始文档：图 4、图 5，

将可见认证图样用于认证文档的一种方式是通过确定文档是原本，即是从原始数字表示形式打印来的，还是非原本，即是从文档影印来的或从非原始数字表示形式打印来的，也就是从对文档的未经授权的数字记录制成的数字表示形式。可以用这种方式来使用可见认证图样的原因是从文档的数字表示形式打印文档并从其数字记录制作文档的数字表示形式或影印

文档总是会导致可见认证图样中信息的损耗，不论打印、数字记录或影印过程多么精确都是如此；因此，可以通过将可见认证图样的原始数字表示形式与通过从文档记录可见认证图样制成的数字表示形式相比较，来确定文档是原本还是非原本。在原始文档的情况下，可见认证图样已经被打印一次并被数字地记录一次；在非原始文档的情况下，可见认证图样已经被打印和数字地记录一次以产生用来制作非原始文档的原始文档，然后依赖于如何制作非原始文档，可见认证图样被影印或者再次被打印和数字地记录，从而在非原始文档的可见认证图样中产生比原始文档的可见认证图样中更大的信息损耗。

基本技术在图 4 中详细示出。在 401 处，示出了利用可见认证图样进行的认证如何对原始文档起作用。文档的原始数字表示形式 403 包含原始可见认证图样 (ovap) 405。然后，在 407 处，原始数字表示形式 403 被打印以产生原始模拟形式 409。打印操作导致模拟形式 409 中的原始模拟可见认证图样 (oavap) 411 中的 $loss1$ 。当认证者 421 认证模拟形式 409 时，它制作 oavap 411 的数字记录，从而导致 $loss2$ 。记录显现为 roavap 415。然后认证者 421 采用比较器 417 来比较 ovap 406 和 roavap 415。其间的差异是 $loss1$ 和 $loss2$ 的和。这在将任何未受其他损毁的 roavap 415 与 ovap 415 相比较时都是成立的，而这样大小的差异是对模拟形式 409 确实是原始模拟形式的可靠指示。

在 420 处，可以看出认证如何对非原始文档起作用。原始文档和非原始文档之间的差异在于非原始文档不是直接从文档的原始数字表示形式 403 打印成的，而是从文档的非原始数字表示形式 423 打印成的，其中文档的非原始数字表示形式 423 是通过数字地记录原始文档 409 而制成的 (422)。作为数字记录的结果，数字表示形式 423 中的非原始可见认证图样 425 经历了额外的信息损耗，这在图 4 中显现为 $loss3$ 。当从数字表示形式 423 打印 (427) 非原始模拟形式 429 时，在非原始模拟可视认证图样中出现另一个损耗，其被表示为 $loss4$ 。当如上所述非原始模拟形式 429 被认证者 421 认证，并且从 noavap 431 制成的 rnoavap 435 被与 ovap 405 相比较时， $loss3$ 和 $loss4$ 的效果将会作为 ovap 405 和 rnoavap 435 之间的比

ovap 405 和 roavap 415 之间的差异更大的差异出现。由于非原始模拟形式 429 中的 noavap 431 将会始终经历额外的 $loss3$ 和 $loss4$ ，因此更大的差异是对非原始文档的可靠指示。

非原始模拟形式 429 当然可以通过影印过程产生，也可以通过记录原始模拟形式（422）以制作非原始数字表示形式 423，然后打印（427）数字表示形式 423 以产生非原始模拟形式 429 的过程产生。获取原始模拟形式 409 的图像然后从图像打印非原始模拟形式 429 的过程导致像 $loss3$ 和 $loss4$ 那样的额外的损耗，因此，以这种方式产生的 rnoavap 435 与 ovap 405 的相似程度仍将比 roavap 425 与 ovap 的相似程度更低。

当然，如果非原始数字表示形式 423 本身是从非原始数字表示形式制成的，则 rnovap 435 也将会包括由对该非原始数字表示形式的影印或者打印和数字记录而导致的额外损耗。显然，如果 $loss1$ 和 $loss2$ 是固定值，则检测器将始终能够确定文档是原始的还是非原始的。但是，一般来说每个损耗都会发生一些变化，例如某些原本比起其他原本来能够以更好的质量（保真度）打印。于是看起来应当采用统计检测方法。

关于区分原始和非原始文档的细节：图 6 和图 7

认证技术的价值实际上就在于其可靠性。使检测差错的概率最小化的关键是用测量从文档记录的可视认证图样与可视认证图样的原始数字表示形式之间的“差异”的方法。所选择的测量方法必须基于受制作非原始文档的过程影响的 VAP 的属性，并且必须清楚地区分原始文档和非原始文档。

我们的方法是将影印、记录和打印过程考虑成滤波器，更具体地说是低通滤波器。从而，比起低频来，高频被打印和记录过程衰减的程度将会更高，并且在每个记录和打印或影印频率将会损耗更多信息。对于其中记录和打印或影印过程保留几乎所有能量的低频，非原始文档中的 VAP 拥有的信息可能不会比原始文档中的 VAP 少很多。很高的频率可能也是没有帮助的，这是因为 VAP 中在这些频率下的大部分能量在 VAP 首次被打印时就损耗了。因此，即使是原始文档的 VAP 也只包含着很少的来自这

些频率的信息。因此，必须对检测器使用的频率进行适当的选择和/或加权。对用于比较的频率的选择以及对用于确定文档是原始还是非原始的阈值的选择一般是通过在来自原始文档的 VAP 上训练比较软件来完成的。

在这里应当指出，上述技术不需要特殊的可视认证图样。相反，整个文档或其一部分可以被用作图样。但是，由于许多文档可能不包含处于确定文档是原本还是拷贝所需的能量级别的信息，因此最好使用包含处于适当能量级别的信息的可视认证图样。在下文中，这种可视认证图样将被称为**拷贝检测图样**（*copy detection pattern*），或 CDP。CDP 中的信息分布在适当的频率中。在优选实施例中，CDP 的原始数字表示形式是通过密钥以伪随机方式生成的，因此能够访问该密钥的程序就随时能够创建 CDP 的原始数字表示形式的新拷贝。该密钥可以被保密，或者只被透露给受信任的当事人。拷贝检测图样被插入或打印在要保护的文档上。在优选实施例中，对来自文档的拷贝检测图样的分析是通过数字地记录文档的 CDP、利用密钥生成 CDP 的原始数字表示形式的新拷贝，并且将记录的 CDP 与 CDP 的原始数字表示形式相比较来完成的。在其他实施例中，可以简单地将记录的 CDP 与 CDP 的原始数字表示形式的先前存在的拷贝相比较。

该技术中使用的算法

本部分描述用于进行以下操作的算法：（1）生成 CDP 的原始数字表示形式；（2）从文档检测和提取 CDP；（3）将 CDP 的原始数字表示形式与记录的 CDP 相比较；以及（4）确定 CDP 是原始的还是非原始的。算法（4）中比较 CDP 的方式以及用于确定 CDP 是原始的还是非原始的阈值是通过训练过程来确定的，在该训练过程中，算法（3）被用于收集训练数据。

生成 CDP 的原始数字表示形式

函数 *make_pattern* 被用于创建可以用从其制作原始文档的数字表示形式的源标识的拷贝检测图样的数字表示形式（*pattern_img*）。*make_pattern* 生成有噪声灰度级或彩色图样。黑边也可以被添加到图样，以帮助在文档

中检测它。CDP 还可以可选地显示标识物。标识物一般将会影响最低频带，因此其对检测的影响将会是有限的。典型值在对参数的说明中给出。

pattern_img = make_pattern(type, height, width, key, filename, border, logo_img, logo_weight)。

用于图样生成的参数

必需：

1. *Type*: 生成的随机数值的类型，例如 “randn”（高斯型 $N(0,1)$ ），“rand”（等概率分布）、“randint”（二元+1 或-1 分布），或 MD5、SHA 算法（0-255 整数）。随机数值随后被用于构成灰度级或彩色图像。
2. *Height*: 图样高度，以像素为单位（例如 104）。
3. *Width*: 图样宽度，以像素为单位（例如 304）。
4. *Key*: 用作随机数生成器的种子的整数值秘密密钥（secret key）或口令。

可选：

5. *Filename*: 保存图样图像的文件的名称。
6. *Registration mark*（例如添加到图样图像边缘的黑边，添加到图样图像四角的点）。
7. *Logo_img*: 要用作背景标识物的图像，自动缩放到图样图像的尺寸。
8. *Logo_weight*: 0 到 1 之间的值，用于对叠加在图样图像上的标识物图像的能量加权（例如 0.2）。

图样生成算法的使用示例：

1. 在特定域中生成图样（例如彩色 RGB 模式中的空间或 DCT 亮度）：

pattern = generate_pattern(type, height, width, key);

2. 如果步骤 1 中的域不是空间域，则将图样变换到空间域（例如逆 DCT）：

```
pattern_img = transform(pattern);
```

3. 如果需要, 将像素值 p 舍入到整数值 $0 < p < 255$ 。

4. 将标识物与图样相结合, 例如混合方程可以如下:

```
pattern_img = (1-logo_weight)*pattern_img + logo_weight*logo_img;
```

5. 添加配准标记 (例如黑边)。

6. 转储图像。

图样图像可以由多个分量/通道构成, 例如红、蓝、绿或 YUV, 它们可以按上述步骤 1 和 2 的方式产生。

为了将 CDP 与标识物或背景图像相结合, 可以采用各种混合函数。例如, 当 CDP 与条形码 (图像) 合并时, CDP 只替换条形码的黑色区域, 而保持白色区域不变。

可以生成任何形状 (例如圆形、椭圆形) 的图样图像。一种简单的方法是使用“形状掩膜”, 其限定由“1”和“0”构成的二维阵列所代表的任意形状。通过将“形状掩膜”应用到矩形图样图像, 可以创建任何形状。

从文档检测和提取 VAP

在本实现方式中, 制作被认证的文档的数字记录, 并且用 VAP 上的黑边来定位数字记录中的 VAP。黑边导致过渡区域的强烈亮度变化, 这是很容易检测到的。也可以使用其他的用于确定 VAP 的位置的技术 (例如文档中的现有特征、黑点等等)。一旦检测到 VAP, 就制作它的数字表示形式, 该数字表示形式可以与 VAP 的原始数字表示形式相比较。数字表示形式是记录的 VAP。

利用以下函数比较 VAP 的原始数字表示形式和记录的 VAP。该函数测量记录的 VAP 与 VAP 的原始数字表示形式的“接近”程度的指标。VAP 的原始数字表示形式可以被存储在检测器的存储器中, 或者如果用于创建原始数字表示形式的参数和函数 `make_pattern(..)` 对于检测器可用, 则 VAP 的原始数字表示形式可以被重新生成。当将图样与标识物相结合时使用的可选参数可能不是必需的, 因为标识物一般只会轻微地影响图样的属

性。用于进行比较的函数是 *analyze_pattern*，其返回 *Results*，并且可以根据实际应用的场景取不同的参数：

Results = analyze_pattern(type, height, width, key, ..., test_img);

或者

Results = analyze_pattern(orig_img, test_img);

参数和输出：

1. *type*、*height*、*width* 和 *key*：这些与对图样生成的说明一致。
2. *test_img*：从文档提取的测试图样图像。
3. *orig_img*：图样的原始数字表示形式。
4. *Results*：包含分析的所有结果。例如，它可以包括为图像的不同元素（例如不同的频率、不同的区域、不同的颜色通道等等）计算的相关性和统计数据的不同度量。

下面的示例示出重新生成原始数字图样的算法的频率以及算法所需的子函数：

1. （可选）从测试 CDP 去除黑边
2. 将测试图样图像变换到它的原始生成域中，例如 8x8 块 DCT：
test_pattern = transform(test_img);
3. 重新生成原始 CDP；
pattern = make_pattern(type, height, width, key);
4. （可选）在本地将测试 CDP 与原始 CDP 同步，如下所述。（可选）向测试 CDP 应用某些图像滤波器（例如锐化），以产生与原始 CDP 的更好的相关性。
5. 如果需要，将原始 CDP 和测试 CDP 转换到进行比较的域（例如 8x8 块 DCT）。注意比较可以在多于一个域中进行，例如既在空间域中进行，又在频域中进行。
6. 为变换后的域中的每个通道计算原始 CDP 和测试 CDP 之间的相似性的若干个测量结果。例如，如果图样是在彩色 RGB 域中生成和记录的，则分析在 8x8 块 DCT 域中进行。于是有 192（即 8x8x3）个组合，通过这些组合两个图样被比较，从而可以执行对相似性的 192 次测量。相似

性度量本身可以用若干种方式来计算，例如通过对值进行分级并且只保留有较高的相关性的值，以便排除测试 CDP 中可能已受破坏的区域。

7. 收集和组合基于其他图像特征的所有一个或多个相似性度量，以测量测试 CDP 的质量或测试 CDP 与原始 CDP 的“接近度”的一个或多个指标。组合函数可以是任何组合不同输入的函数，例如通过向原始 CDP 和测试 CDP 之间更有区分力的特征分配更高的权重或重要性来组合相似性度量的函数。

如上所述，复制过程总是会有损原始 CDP，并且一般预期对于从模拟形式记录的 CDP，接近度或质量的不同度量会更低。但是，由于统计变化，在确定测试 CDP 是从原始模拟形式记录的还是从非原始模拟形式记录的时，对不同度量的充分选择和组合将会更有效。

图 6 示出对于 30 个频带（在 603 处示出），来自被认证的文档的原始 CDP 和测试 CDP 中的频率的能量之间的相关性（在 605 处示出）。正如预期，能量之间的相关性在低频带中最高，而在高频带中最低，其中，在拷贝过程中从低频带只损耗很少的信息，而在高频带中，即使是单个打印操作也会导致大部分信息的损耗。如果在中频带中的相关性比来自原始文档的 CDP 的平均值低得多，则 CDP 不是原本，因此被认证的文档也不是原本。这是图 6 的图线的情况，该图因而示出了被认证的文档不是原本。

当相关性值本身不足以确定文档是原始模拟形式还是非原始模拟形式时，也可以考虑其他图像特征。可以用于产生原始 CDP 和测试 CDP 之间的相关性值的额外的图像特征包括：

- 颜色直方图
- 边缘、线条和轮廓
- 其他域（例如傅立叶和小波域）中的频率
- 亮度和对比度

检测 CDP 是来自原始文档还是非原始文档

函数 *detect_pattern* 分析由 *analyze_pattern* 返回的结果，并且返回值 *Output*，该值指示 CDP 是来自原始文档还是非原始文档。

Output = *detect_pattern*(*Results*, *Parameters*)

Results: 可以是标量值或向量, 函数 *analyze_pattern* 的输出。

Parameters: 调节检测函数的行为所需的值, 它可以依赖于应用的要求和执行检测的条件。

Output: 可能有不同的输出值。就其最简单的形式而言, *Output* 可以取三个值: ORIGINAL、NON-ORIGINAL 或 PROCESSING-ERROR。最后的输出可能发生在图样被恶劣地记录时。*Output* 可以返回更详细的信息, 例如 NON-ORIGINAL 可以进一步指示来自非原始文档的测试图样是如何产生的(例如复制、影印、重新生成等等)。*Output* 还可以提供质量或接近度的指标。

以下是简单检测函数的算法的示例:

1. 组合由 *analyze_pattern* 返回的各种 *Results* 值, 以获得标量值 *S*。做到这一点的一种方式将会是通过对返回的 *Results* 求和来产生 *S*。

2. 如果 $S > T1$, 则输出 ORIGINAL, 否则如果 $S > T2$ 则输出为 NON_ORIGINAL, 否则输出为 PROCESSING ERROR。

在这里 *T1* 和 *T2* 是两个标量参数, 它们一般是经由训练过程获得的, 其中一般 $T1 > T2$ 。

来自文档的 CDP 与原始 CDP 的本地重同步

为了将从文档记录的 CDP 与原始 CDP 相比较, 记录的 CDP 必须与原始 CDP 同步。做到这一点的一种方式是使用记录的 CDP 中的同步点, 例如黑边 601, 来同步原本。一旦 CDP 被同步, 其间的比较就被逐像素或逐块地进行。

当在文档中打印 CDP 时或在从文档数字地记录 CDP 时发生差错时, 就不能用这个方法完全地同步 CDP。例如, 在原始 CDP 和从文档记录 CDP 之间必然有少于一个像素的移位。此外, 在图样上移位可能发生变化: 在某些情况下, 记录的 CDP 的上半部分与原始 CDP 相比可能向下移位, 而下半部分向上移位(当然也可能反之)。这些移位可能很难注意

到，可能不会一贯地发生，或者可能在记录的图样中发生本地变化。它们一般是由打印机中的轻微不稳定性导致的，但是也可能由记录设备中的类似的不稳定性导致。

这些不可预测的亚像素移位可能降低检测器的性能：由于这些失准（misalignment），某些来自原始文档的 CDP 可能被检测成来自非原始文档的。一种应对这些来自原始文档的“病态”CDP 的方法以及一般来说改进 CDP 检测的稳定性的方法是在本地对 CDP 进行重同步，以校正本地失准。存在若干种执行本地重同步的方式，但是一般概念是利用记录的 CDP 本身来进行本地重同步。

一种执行本地重同步的方式是将原始 CDP 划分成块（不交迭的块是优选的，但是块也可以交迭），并且找出记录的 CDP 中的哪个块与原始 CDP 的给定块具有最接近的匹配。如果没有失准，则记录的 CDP 的与给定块最接近匹配的块在记录的 CDP 中的位置将会与给定块在原始 CDP 中的位置相同：例如原始 CDP 的具有开始位置(80,80)和结束位置(89,89)的 10x10 块的最佳匹配将会是记录的 CDP 的相应的块(80,80)到(89,89)。但是，如果存在失准，则最佳匹配也可能是与块(81,80)到(90,89)（向右移位一个像素）。如果是这样的话，则记录的图样将会使块(81,80)到(90,89)向左移位 1 个像素到位置(80,80)到(89,89)。同样的概念可以应用到记录的 CDP 中的每个块，以产生“经本地重同步的”CDP。

本地重同步要求几个参数和函数。首先，我们必须定义原始 CDP 的每个块和记录的 CDP 的具有相同尺寸的块之间的距离的度量。用于此目的的方便的度量是标准相关系数。还有必要设置原始 CDP 被划分成的块的尺寸；一般可以使用尺寸为 8x8 或 16x16 的块，但是一般来说可以使用尺寸为 $N \times M$ 的块。如前所述，块可以是交迭的，在这种情况下，需要定义连续的块之间的交迭量。要设置的另一个参数是搜索范围或搜索区域：从匹配位置开始，算法为了搜索匹配块应当进行到何种程度？这是用参数 n 来设置的，其中对于开始于原始 CDP 的位置 (x,y) 的块，所有位置为 $(x+/-i, y+/-i)$ ($0 < i < n$) 的块都被测试。

还可以在本地重同步之前对数字和记录的 CDP 进行缩放：这允许

了颗粒更细的匹配。例如，通过将两个 CDP 放大 2，可以恢复半像素移位。最后，可以对经重同步的 CDP 迭代应用同步算法，直到不再发现进一步的改善。

一旦执行了重同步，就可以执行经重同步的记录的 CDP 和原始 CDP 之间的相似性/距离的任意度量。可以执行简单的相关或本地频率分析，其中可能利用基于训练集合的参数。这些度量一般在整个 CDP 上对某个量取平均，但是对于可能发生在某些应用中的对扫描的 CDP 的某些本地损毁并不是始终鲁棒。例如，在某些情况下，CDP 的一个区域可能是恶劣地打印的，或者可能受到了刮擦、书写或水的损毁。在其他情况下，扫描设备可能向扫描的 CDP 插入了失真，这个问题一般发生在馈通型设备上，并且是在文档未被正确插入时发生的。为了使得 CDP 对于这些种类的失真更为鲁棒，可以使用相似性的更鲁棒的度量：一个这种度量是中值本地相关系数，其中为 CDP 的每个块计算一个相关系数，并且计算所有本地相关系数的中值。在这里，计算中值而不是平均值使得检测器对于本地更改明显更加鲁棒。为了对付 CDP 中的大量经破坏的区域，还可以只计算能够被假定为非经破坏的 20%的最佳本地相关系数的平均值。在一种实现方式中，计算此类“有偏”平均值的这个计算过程被单独应用到每个频率通道，并且可选地应用到不同的颜色通道。当然，前述同步技术不仅可以用于 CDP，还可以用于任何记录的需要与原始可见认证图样同步的可见认证图样。

CDP 的应用

CDP 可以用于任何它有助于区分原始文档和非原始文档的情形。CDP 可以通过任何以充分的保真度打印 CDP 的过程来打印，以便 CDP 的数字记录能够与 CDP 的原始数字表示形式相比较。图样可能尤其适合于检测由特定的影印、扫描或打印技术制成的非原始文档。CDP 的特殊使用包括：

1. 将 CDP 打印在包装中以进行商标保护
2. 将 CDP 打印在支票或货币上以进行拷贝检测
3. 将 CDP 打印在包括证书、合同之类的贵重文档上，以验证文档是原本还是拷贝。

4. 将 CDP 打印在全息图上

5. 将 CDP 打印在贵重货物上，例如航空/汽车零件或药物。

更一般而言，CDP 可以用于任何需要能够确定向文档应用了什么过程的应用中。图样当然可以根据需要而改变以最佳地检测感兴趣的过程。

CDP 也可用于下面的应用：

1. 打印质量的基准检查

当读取 CDP 时，CDP 的数字记录的质量指标被计算。该质量指标将会根据打印质量、纸张/基底质量或数字化/扫描（设备）质量而改变。CDP 质量指标于是可以用于量化某个打印过程、某个衬底或某个扫描仪的质量。

2. 质量控制

同样，CDP 读取器可以被用在打印生产过程中，用于自动质量控制。CDP 与手动检查相比的优点在于它给出了对质量的自动、客观且准确的度量。

3. 跟踪

CDP 具有与打印机、纸张、照相机和使用和磨损相关联的结构和特性。原则上，对 CDP 的分析可以确定文档的一般“历史”：它是如何被打印的，以及它经历了什么样的“磨损”。

VAP 的实现细节

文档中 VAP 的形式

用 VAP 来检测模拟形式中的更改所需的只是模拟形式中存在着具有会达到此目的的图样的区域并且该图样的原始数字表示形式能够与从模拟形式记录的图样相比较。因此，在某些情况下可以将模拟形式中的先前存在的图样用于该技术。虽然更通常的情况是 VAP 将会作为新模拟形式的设计的一部分被包括。当然不需要将 VAP 隐藏在模拟形式中，并且在某些情况下，其实可以宣传其存在以向消费者保证可以非法模拟形式能够被检测到。另一方面，VAP 可以具有任何形状，从而可以很容易被构建到模拟形式的其他特征中。图 8 示出两个示例。在 801 处示出了条形码，其条

形构成了 VAP。在 803 处是包含 VAP 的标识物。当然，在文档中可以有多多个 VAP，并且多个 VAP 可以共享一个位置。这可以通过向每个图样赋予加权值以使所有图样的权重加起来为 1 来完成，例如：

$$Final_pattern = a * pattern1 + (1-a) * pattern2, \text{ 其中 } 0 < a < 1$$

多个图样的一个应用是合同的认证，其中每个当事人在签署合同或以其他方式终止协商中的一个阶段时添加其自己的图样。

还可以将一般以不同密钥产生的若干个 CDP 插入到文档上的不同位置，以使得多个当事人能够验证其自己的 CDP，而无需能够验证其他当事人的 CDP（因此无需能够复制它们）。甚至还可以利用不同的密钥来生成 CDP（每个密钥可以控制 CDP 的不同空间区域和频率区域），以使得不同当事人能够验证 CDP。这样，如果一个当事人发布其密钥，这个密钥不足以对 CDP 进行原样复制（所有密钥都是必需的），并且安全性不会受到危害。这与“共享秘密”的概率类似。

VAP 的配准

优选实施例采用黑盒 106 作为 VAP 的配准。但是，也可以采用许多其他配准技术。例如，可以使用已经显示在包装上的诸如边框、条形码之类的可见图样来定位 VAP，以及 OCR。还可以使用 UV 标记或 8/24/04 授权的 Zhao 等人的标题为“*Apparatus and methods for improving detection of watermarks in content that undergone a lossy transformation*”的美国专利 6,782,2116 中描述的任何技术。此外，还可以对记录的 VAP 进行傅立叶-梅林变换，并将其与 VAP 的原始数字表示形式相匹配。

对于某些应用，难以得知 VAP 的数字记录的方向是否正确，或者在读取前它是否应当被颠倒过来（180 度旋转）。为了避免必须分析 VAP 一次，然后如果分析不成功又必须在相反的垂直方向上旋转它并再次分析它，可以设计对称 VAP：下半部分是上半部分的镜像。于是可以独立于其垂直方向来分析 VAP。

VAP 的图样的属性

图样可以是灰度级图样或者可以是彩色图样。在后一情况下，可以采用不同的颜色通道，例如 RGB 和 YUV。图样也可以在各种频率域中生成，例如空间、小波、DFT 或 DCT 域。

生成 VAP

VAP 的噪声性（即随机性）正是使得造假者和伪造者难以应对的性质。任何能够产生随机或伪随机图样的技术都可以用来生成 VAP。在优选实施例中，是通过向伪随机数生成器提供值来完成生成的，该伪随机数生成器生成对该值为说唯一的随机数序列。从而该值充当可用于生成图样的新拷贝的密钥。在不同的实施例中可以使用不同的伪随机数生成器，并且可以从不同的概率分布中取得生成随机数的概率频率值。密钥也可以用于确定在 VAP 中执行分析的位置。正如下文中对利用 VAP 来携带其他信息的论述中所说明的，密钥可以包括这种其他信息。在某些应用中，用于设计图样的密钥可能不被透露给其他当事人。在这种情况下，可以使用任何有用的分发密钥的方式，例如非对称密钥或公钥-私钥对。

通过将标识物添加到图样或者反之，可以使图样与标识物相结合。标识物可以是任何现有的图像或文档，包括用于其他目的的图像（2-D 条形码，带水印的图像，等等）。还可以按标识物将会最低限度地干扰记录的 VAP 与 VAP 的原始数字表示形式的比较的方式向图样或标识物应用例如滤波之类的任何过程。

打印 VAP

由 VAP 提供的认证的质量完全依赖于 VAP 被打印在文档上时的保真度。如果在打印过程结束时添加了“质量控制”步骤以保证 VAP 的保真度，则可以减少认证差错：

1. 每个打印的 VAP 将会被传递到自动验证过程，以检查认证图样是否具有将其识别为原本所需的最低质量。
2. 如果质量低于最低质量，则将发出警告，并且将重新打印包含认证图样的文档/包装。

3. 这种验证也可以充当用于由打印机引起的打印质量或差错的“质量控制”。

VAP 的生成可以适应于打印技术。例如，如果采用只打印二元点的激光打印机，则可以生成二元点 VAP 以更好地利用打印机的潜力。此外，在打印机的颜色空间中，VAP 可能被更充分地生成和打印。如果某个打印机使用特定的墨水（例如 CMYK），则在该域中生成 VAP 将比在 RGB 域中生成 VAP 更有效。如果 VAP 是用只能产生二元点的激光雕刻机刻在金属中的，则生成二元 VAP 将会更有意义。

利用 VAP 来携带其他信息

下面论述三种利用 VAP 来携带其他信息的方法：预留 VAP 的某些区域以保存信息，利用其他信息来生成用于制作原始 VAP 的密钥，以及向 VAP 添加水印。添加水印的不利之处在于它减小了 VAP 检测非原始模拟形式或 VAP 中的修改的能力。

预留 VAP 中的区域以保存信息

VAP 的某些区域（例如 8x8 块）可以被预留以用于保存信息。在某些区域中，VAP 的结构/特性实际上并不用于验证其真实性，而是用于存储某些信息比特。可以利用密钥来伪随机地选择这些区域，以便没有密钥的实体不能确定 VAP 中的区域实际上是用于存储信息还是用于确定 VAP 的真实性。在用于保存信息的区域中，VAP 的某个结构/特性可以对应于信息的某个比特值（“0”或“1”）。这些依赖于比特的结构/特性当然可以根据密钥所确定的而改变。注意，预留的区域和它们包含的信息是生成的 VAP 的一部分。从而它们不会降低 VAP 检测非真实文档的能力。预留的区域的一种用途是存储用于生成 VAP 的密钥。

利用信息来生成 VAP 的密钥

本论述使用以下术语：VAP 是用密钥 P 创建和检测的；如上文针对预留区域或下文针对水印所描述的，可能希望使用不同密钥 S 来将消息嵌入

到图样中；消息 M 被用密钥 S 嵌入在 VAP 中；最后，附加的信息 I 可以按可见的方式打印在文档上（序列号、条形码等）图样之内或之外，或以不可见地方式被 UV 编码在图样之内或之外，或者从外部源获得。

固定图样密钥

在一个实施例中，VAP 创建密钥是固定 P 。标准偏移打印技术通常就是这种情况，其中打印技术没有为每个封装/产品/文档动态地改变图样的能力。密钥可以如上所述被保密，或者可以被结合到其他安全性特征中。例如，可以用 UV 墨水将其打印在文档上。固定图样密钥一般可以用于商标保护或者文档保护。

可变图样密钥 在另一个实施例中，VAP 的密钥依赖于秘密密钥 S 和某个其他信息 I 。这个其他信息 I 可以被显示在文档上（在图样之内或之外），或者从外部源获得。来自文档的信息例如可以是序列号、文本、条形码等。来自外部源的信息例如可以是一个值，该值与 VAP 相关联，并且为检查包含 VAP 的文档是否真实的人所知。图样密钥可以是秘密密钥和信息 I 这两个参数的任何任意函数 $P=f(S,I)$ 。一个简单的函数是对两个参数的连结或求和，但是许多其他函数也是可能的，例如两个参数的组合的散列值，等等。在检测时，利用适当的技术 – 条形码读取器、OCR 等等 – 提取打印的信息 I 。然后，按 $P=f(S,I)$ 生成图样密钥。典型用途包括利用数字打印的商标保护。

VAP 中的水印

可以利用任何水印技术将可见的或不可见的水印嵌入在 VAP 中。水印可以用于多个目的。它可以包含任何信息，其中包括只是单个比特，如上所述，或者帮助图样的配准。可以利用用于生成 VAP 的密钥或利用其读取局限于另一用户或用户群组的另一个密钥来检测水印。在下文说明的第三种可能是利用水印携带的消息来导出用于生成 VAP 的密钥。

当数字水印被嵌入到 VAP 中时，VAP 将会被轻微地修改。结果，当

同样的 VAP 被用于真实性验证时，它就该目的而言的可靠性将会降低。作为替换，数字水印可以被嵌入到 VAP 中如上所述的预留用于存储信息区域中。

水印和密钥

在另一个实施例中，图样创建密钥 P 是从秘密密钥 S 和作为数字水印嵌入在拷贝检测图样中的消息 M 导出的。在这种情况下， M 取代上述用于创建可变图样密钥的信息 I 。在创建时，图样密钥 P 可以是秘密密钥 S 和消息 M 的任何函数 $g(M, S)$ 。以通常方式生成图样，然后将水印插入到图样中，其中水印利用秘密密钥 S 作为参数来对消息 M 进行编码。在检测时，首先必须利用秘密密钥 S 从图样中读取水印消息 M 。一旦 M 已知，图样密钥 $P=g(M, S)$ 就被导出，并且图样被分析。

在这个应用框架中，不需要用辅助技术来提取打印在包装上的更多信息。但是，可以在这里描述的原理内以若干种方式使用打印在包装上的信息 I 。例如，秘密密钥 S 可以与信息 I 结合使用，以产生水印密钥 W ，即 $h(S, I)=W$ ，它被用于将消息嵌入在图样中。然后，以与先前相同的方式生成图样密钥， $P=f(M, W)=f(F, h(S, I))$ 。一般来说，VAP 可以与水印技术和其他读取技术（例如 OCR 或条形码读取器）相结合，以产生不同级别的验证。

比较 VAP

记录的 VAP 如何与 VAP 的原始数字表示形式相比较将会依赖于 VAP 是如何制成的以及它的用途是什么。某些普遍适用的变化包括独立地评价某个区域，以便拥有关于应用到该文档的过程的更多线索，或者找出安全性特征。如上所述，VAP 可以包含不止一个认证图样，并且不同的图样可以被不同的群组分析。

在 VAP 能够被有意义地比较之前，比较程序必须被用从原始文档记录的 VAP “训练”，如上文针对 CDP 所述。训练确立用于确定从其真实性正在受到检查的文档记录的 VAP 是否真实的阈值。阈值的意义当然将

会依赖于用 VAP 来检测的更改的种类。每当打印原始文档的方式以影响 VAP 比较的方式发生变化，就需要重新训练。可以通过 p 将多个 VAP 打印在一张纸上，扫描这张纸，并且将扫描结果提供给训练软件，来自动地完成训练。

在另一个实施例中，不是将测试 VAP 的数字记录与相应的数字表示形式相比较以测量其质量指示，而是可以将数字记录与另一个 VAP 的数字记录（一般是被扫描的原始 VAP）相比较。

执行 VAP 分析的环境

进行 VAP 分析所需要的是能够从文档记录 VAP 以制作记录的 VAP 的设备，VAP 的原始数字表示形式的拷贝，以及能够将记录的 VAP 与 VAP 的原始数字表示形式相比较的处理器。记录器和处理器可以在彼此本地，或者可以通过网络连接。网络可以是局域网（LAN）或广域网（WAN）。本地环境的示例是具有扫描仪的 PC、分析代码的拷贝以及 VAP 的原始数字表示形式的拷贝。VAP 的原始数字表示形式的拷贝可以是下载的、预先存储在本地的或者利用密钥在本地生成的。分析结果被输出到 PC 的显示设备。

在网络环境中，扫描、分析和 VAP 的原始数字表示形式可以按任何方式在网络上分发。维护 VAP 的原始数字表示形式的安全性并简化本地级别上所需的设备的分发是这样的分发：在这种分发中，扫描是在连接到 WAN 的设备中完成的。当文档上的 VAP 已经被扫描以产生记录的 VAP 时，记录的 VAP 被发送到 WAN 中的一个位置，在该位置处分析代码和 VAP 的原始数字表示形式两者都是可用的。原始数字表示形式可以被存储或者可以根据需要重新生成。分析在该位置处完成，并且只有分析结果经由 WAN 被返回到用于扫描的设备。在网络环境中，一般来说，记录的 VAP 中携带的信息或与记录的 VAP 一起发送的信息可被用于取得供分析时使用的信息。例如，文档可以包含序列号，并且序列号可以与记录的 VAP 一起被发送到进行分析的位置。如果在 VAP 和序列号之间存在关联，则序列号可以被应用到该位置处的数据库或网络中的其他位置，以取

得用于应当与记录的 VAP 相比较的 VAP 的原始数字表示形式的密钥或 VAP 本身的原始数字表示形式的拷贝。如上所述，序列号可以作为 VAP 中的可见水印在包含 VAP 的条形码中指定，可以是来自文档 OCR 来的，或者甚至可以由进行扫描的人输入的。

照相机（网络照相机、便携式摄像机等）也可以用于捕捉 VAP 的图像。在这种情况下，VAP 检测器不仅接收一个图像作为输入，还接收恒定的图像流作为输入。由若干个图像提供的额外的信息在分析时可能是非常有用的。但是，由于分析一个图像所需的时间可能远大于两个连续的图像之间的时间，因此图像流的使用可以被优化。例如，可以从流中选择看起来具有用于正确读取的属性（良好的锐度、VAP 完全包含在图片中）的图像，并将其用于分析。

将 VAP 与其他安全性技术相结合

VAP 可以与旨在使模拟形式更安全的其他技术相结合。例如，VAP 可以与例如数字水印这样的信息隐藏技术、例如 1-D 或 2-D 条形码这样的机器可读信息、全息图或者能够应用到模拟形式的任何其他技术一起使用。技术之间的关系可能是各种各样的：例如，2-D 条形码可以包含独立的信息，或者图样分析所需的秘密密钥，或者反之，VAP 可以保存对 2-D 条形码进行解码所需的密钥，或者 2-D 条形码可以包含 VAP。

检测 VAP 的位置

当模拟形式被扫描时，在扫描开始时不是总有可能得知 VAP 位于何处。这可能是因为使用 VAP 的应用必须支持具有不同格式的文档和/或放置在不同位置的 VAP，因为将文档放置在扫描仪上的最终用户不知道如何将文档放置在其上；配备有照相机的便携式照相机/计算机被用于捕捉文档的图像，而在图像捕捉时存在固有的人类变异性；或者就只是因为每次扫描中存在天然的变异性，并且文档中靠近 VAP 的图样干扰了 VAP。在一个“最坏情形应用”中，letter 大小的模拟形式的整个区域被扫描，并且 VAP 可能位于模拟形式上的任何位置并且可能具有任何方向。

但是，可以利用 VAP 的一般统计属性来在模拟形式中定位它。区分 VAP 与大多数其他图像或文档特征的属性是其直方图的分布。由于 VAP 是有噪声的，因此原始数字 VAP 的像素值（对于灰度级图像有 256 个；对于二元图像有 2 个；对于彩色图像有 24 位或更多）中的每一个一般是等概率的，或者可以具有特定的分布。虽然 VAP 的打印和扫描确实会修改其像素值的分布，但是这样产生的修改是非常特殊的。通过在校准过程中打印和扫描一定数目的 VAP，可以制作一个直方图，该图是对从模拟形式扫描 VAP 中的像素值的平均分布的估计。图 13 示出了一个这种分布 1301，在下文中它被称为“模板直方图”。由于 VAP 的有噪声性质，模板直方图中像素的分布一般将会比文档中像素的分布宽得多。

文档的扫描图像可以被划分成块，这些块的大小一般是 50x50 像素。对于以 300dpi 扫描的 letter 大小的扫描仪，3300x2550 像素的扫描图像被划分成 $66 \times 51 = 3366$ 个块。这 3366 个块中每一个的直方图被计算，并且被与模板直方图相关。正如从文档的大多数部分的非随机性可以预期的，已经根据经验观察到扫描的模拟形式的图像中的大多数块的直方图与模板直方图的相关性接近零，而 VAP 的块与模板直方图的相关性就明显是正的。与模板直方图的相关性最高的块一般可以被假定为属于 VAP，尤其在相邻块也显示高相关性的情况下更是如此。本地搜索算法可以被应用来检测属于该 VAP 的所有相邻块，然后包含 VAP 的区域可被剪切下来并被提供作为恢复函数的输入。

上述方法要求拥有关于给定打印-扫描环境的足够的知识来产生模板直方图。但不是总有可能拥有这种知识；在这种情况下，VAP 的特有的高平均信息量可以被用来检测它。这里所使用的平均信息量 (*entropy*) 是灰度级图像的块中的像素具有大量不同值之一的概率。例如，在包含印迹（打印的文本或图形）的块中，像素一般将会集中在两个值附近，即，如果代表纸则接近白，如果代表印迹则接近黑，从而像素会具有大量值之一的概率就较低，因此平均信息量也较低。因为数字表示形式中的 VAP 是有噪声的，因此 VAP 中的像素会具有大量值之一的概率较高，从而 VAP 的平均信息量较高。虽然数字表示形式中的 VAP 的平均信息量由于打印-扫描

而减小，但通常具有 VAP 的区域仍保持为扫描图像中具有最高平均信息量的区域。因此，通过测量扫描图像的每个区域中的平均信息量并选择具有最高平均信息量的区域，可以在不使用模板直方图的情况下导出 VAP 的位置。

在某些情况下，这个方法不起作用，因为包含 VAP 的数字表示形式包括其他含丰富纹理的区域，在这些区域中，在打印-扫描期间纹理被很好地保持，因此平均信息量被很好地保持。在这种情况下，避免检测这种区域的一种方式是对进入平均信息量计算的那组可能的像素值设置某些限制。例如，如果在扫描图像中 VAP 通常具有 0 到 150 之间的值，则可以从平均信息量计算中排除所有亮度高于 150 的值。当然可以具体地将 VAP 设计为具有与带纹理的区域中的值范围不同的值范围。

VAP 的其他属性，例如其大小、其相对于其他安全性特征的位置或它是以特定墨水打印的这一属性，可以被用作额外的因素，以进一步区分 VAP 和非 VAP 区域。

这种方法对于定位较低质量的拷贝或伪造品可能不起作用，这是因为 VAP 属性可能严重受扰。在这种情况下，另一个方法是利用 VAP 一般比图像的任何其他部分都大得多的动态范围（dynamics），即使它被拷贝也是如此。为了测量一个区域（例如 50x50 像素的区域）的“动态范围”，可以测量像素和相邻的每个像素之间的平均差异。

对于所使用的任何方法，一旦一个块被识别为是 VAP 的一部分，则仍有必要在该块周围进行搜索以找出属于 VAP 的所有其他相邻块。任何本地搜索算法都可用于找出具有给定属性的一组相连的块，其中所述属性是该块对于上述函数具有“显著”输出。

利用 CDP 来检测非原始数字表示形式和非原始模拟形式 图 10 和图 11

打印和扫描的效果可以被视为本质上对应于将一定量的随机噪声添加到文档和文档的 CDP 的每个像素值。由于模拟形式总是在其被打印和扫描之后被评价，因此预期打印和扫描过程向非原始模拟形式添加了比原始模

拟形式更多的噪声。因此，一般可以通过测量 CDP 的数字原本和来自扫描的模拟形式的 CDP 之间在空间域中的简单差异来区分原始模拟形式和非原始模拟形式。如果 CDP 是来自非原始模拟形式的，则将预期较高的距离。位于第 i 列第 j 行的像素 $x(i,j)$ 和 $y(i,j)$ 分别是数字原本和测试图样的像素值，数字原本和测试图样之间的距离度量为：

$$D = \sum_j \sum_i |x(I,j) - y(I,j)|^p / (N * M)$$

其中 p 是任意正数， N 和 M 是图样上的以像素为单位的宽度和高度。如上所述，原本的距离 D 始终会高于拷贝的。

虽然可以从数学上证明上述度量对于区分拷贝和原本几乎是最优的，但是如果不对来自模拟形式的 CDP 进行某些处理，这就不能适用。其原因是打印和扫描导致模拟形式中的像素值的非线性变换。这些变换随着打印机和扫描仪而发生变化，甚至随着打印/扫描参数而发生变化。一般来说，来自模拟形式的 CDP 像素值的频谱将会被打印扫描序列压缩，使得来自模拟形式的 CDP 的频谱的极值比来自原始数字表示形式的 CDP 中的极值稀少得多。

但是，这个问题可以通过向来自模拟形式的 CDP 应用直方图转换函数来克服。直方图转换函数修改来自模拟形式的 CDP 的每个像素，从而使得直方图等同于来自原始数字表示形式的 CDP 的直方图。这个变换函数一般是在校准步骤中利用一组打印的 CDP 来估计的。变换函数一般是固定的，并且被应用到每个来自模拟形式的 CDP。图 10 在 1001 处示出了示例性直方图转换函数。图 11 示出了向来自模拟形式的 CDP 应用直方图函数的效果。来自原始数字表示形式的 CDP 在 1101 处示出；从模拟形式扫描的 CDP 在 1103 处示出；通过应用变换函数而校正的 CDP 1103 在 1105 处示出。从图 11 明显可见，经校正的 CDP 1105 比起 CDP 1103 来更接近来自原始数字表示形式的 CDP 1101。

模拟形式中的 CDP 的平均亮度一般具有某些变化，这是由于扫描仪中的不同照明条件和/或喷射到纸中的不同墨量。这种变异性是自然发生的，无法被控制。为了使这种变异性的效果最小化，可以确定来自原始数字表示形式的 CDP 的平均亮度，确定扫描的 CDP 的实际亮度，然后向扫描的

CDP 的每个像素添加固定值或从中减去固定值，从而使得其平均亮度等于来自原始数字表示形式的 CDP 的平均亮度。例如，如果来自原始数字表示形式的 CDP 中的像素的平均值是 127，而扫描的 CDP 中的像素亮度的平均值实际上是 118，则向扫描的 CDP 中的每个像素加上 9。这种像素亮度调节一般是在直方图转换之前应用的。

在这种变换被应用之后，来自模拟形式的 CDP 的像素值将会与来自原始数字表示形式的 CDP 的像素值具有相的频谱。因此它们是可比较的，并且方程 1 可以适用。如果 $f()$ 是直方图转换函数，则距离 D 由下式给出：

$$D = \sum_j \sum_i |x(I,j) - f(y(I,j))|^p / (N * M)$$

注意：此距离函数只是一个示例；也可以使用若干个其他距离函数。

例如，可以向数字表示形式而不是模拟表示形式应用变换，即：

$$D2 = \sum_j \sum_i |g(x(I,j)) - y(I,j)|^p / (N * M)$$

应对打印-扫描环境的变化

在某些应用中，在用于打印模拟拷贝的打印-扫描环境中可能有大量变化。发生变化的原因包括以下：

- 包含 CDP 的模拟形式是在不同打印机上打印的；
- 模拟形式是打印在不同纸张或基底上的，或者不同的层被覆盖在 CDP 上。
- 不同的扫描仪被用于扫描模拟形式。
- 对模拟形式的不同物理处理。

在不同模拟形式的 CDP 中产生这种变化的应用的一个示例是在由具有不同属性的不同纸张制成的不同类型的信封上打印 CDP，有时甚至在不同的打印机上打印 CDP 的邮资计算器。

在这种应用中，直方图转换函数可能受上述属性的每个组合的影响。应用错误的直方图转换函数一般将会来自模拟形式的 CDP 和来自原始数字表示形式的 CDP 之间的距离的失真。于是，一种解决方案是使用若干个直方图转换函数，并且在验证来自模拟形式的 CDP 时，将每个直方图转换函数应用到 CDP，找出被直方图转换函数修改的与来自原始数字表示形式的

CDP 具有最短距离的来自模拟形式的 CDP，并且只将这个修改的 CDP 与来自原始数字表示形式的 CDP 相比较。如果其他参数依赖于特定的打印机-纸张-扫描仪组合（例如阈值），则对“最佳”直方图转换函数的选择将会意味着对其他参数的选择。

也可以用其他方式来使不稳定的打印-扫描环境的影响最小化。考虑用于一个应用的一组可能的直方图转换函数，其中可以为每个来自模拟形式的 CDP 生成一个转换函数，则存在不同的方式来综合自然的变化。一个解决方案包括测量每个像素的标准偏差，然后利用它作为在测量来自模拟形式的 CDP 和来自原始数字表示形式的 CDP 之间的距离时考虑的归一化因子。另一个解决方案包括估计每个像素亮度的典型值的上界和下界（例如亮度 100 在扫描图像中应当具有 90 到 110 之间的典型值，而亮度 40 可能具有 20 到 60 之间的典型值，其范围是两倍大），并且更重地处罚位于这些边界之外的像素。另一种解决方案包括具有忠诚地代表应用中可能发生的不同转换函数的频谱的一组不同的转换函数，并且每次都使用最适当的那个，从而产生来自模拟形式的 CDP 和来自原始数字表示形式的 CDP 之间的最小距离。

甚至更保守的解决方案也是可能的。一种解决方案是使用预先估计的参数，但是在来自模拟形式的 CDP 上估计它们。这允许了对变化的更大的容限，但是这种方法有两个潜在的局限性：

- 由于参数是在来自模拟形式的 CDP 上估计的，因此这允许了模拟形式的更大的容限，而这种模拟形式不重视来自原始数字表示形式的 CDP 的典型直方图，因此可能是非原始模拟形式。
- 由于参数是只从一个来自模拟形式的 CDP 估计的，因此可以预期对参数的估计不那么准确。例如，对于具有 255 个等概率的亮度值的 10000 个像素 CDP，对于每个亮度值平均将会有少于 40 个样本，并且由于自然统计变化，某些亮度值可能具有远少于 40 个样本。

可以按不同方式来对待第一问题。一种方式是像标准方法中那样拥有预先估计的转换函数，但是不是用它来转换来自模拟形式的 CDP，而只是用来测量与为 CDP 模拟形式估计的转换函数的距离。转换函数之间的距离

的一种可能的度量为：

$$D(f', f) = 1/256 * \text{Sum}(\text{lum}=0 \text{ to } \text{lum}=255) \text{ abs}(f'(\text{lum}) - f(\text{lum}))$$

其中 $f'()$ 和 $f()$ 分别是自估计的转换函数和平均转换函数。还可以输入与每个像素的自然变化相对应的归一化因子 $g(\text{lum})$ ，例如：

$$D_n(f', f) = 1/256 * \text{Sum}(\text{lum}=0 \text{ to } \text{lum}=255) \text{ abs}(f'(\text{lum}) - f(\text{lum})) / g(\text{lum})$$

该距离可以用作可以进入判决的额外的证据。例如，在利用自估计参数转换计算与来自原始数字表示形式的 CDP 的距离时，两个不同的来自模拟形式的 CDP 可能具有相同的相似度 78。但是，它们的自估计转换函数可能与平均转换函数有不同的距离，例如对于第一 CDP， $D_n(f', f) = 2.5$ ，对于第二 CDP， $D(f', f) = 0.5$ 。第一 CDP 的较大的距离可用于确定它是非原始模拟形式（虽然其质量较高），而对于相同的质量指标，第二 CDP 将会被认为是原始模拟形式。

通过假定转换函数的模型，例如转换函数遵循某个回归函数，例如多项式回归函数或对数回归函数，可以应对第二个问题。这使得要估计的参数的数目最小化，并且给出了没有不连续点的更平滑的函数。

最后，在某些情况下，打印-扫描属性可能随时间而演变，以及/或者最初不可能校准 CDP 检测器。在这种情况下，可行的方法包括结合每次新扫描的数据，在开始不知道环境参数时允许更大的容限，并且逐渐减小该容限，这是因为新数据的添加允许了对基本参数的更准确估计。该方法在本质上是通用的，并且对于与 CDP 检测有关的所有参数（转换函数、阈值等）都是有效的。为了结合新的知识，在分配给先验指示的重要性逐渐减小的情况下可以应用贝叶斯学习。通过估计获得的信息可以被存储在数据库中，或者在不同验证站之间共享。这种方法允许了将与打印时的 CDP 质量相关的信息与扫描参数解除耦合。信息是在发生验证时估计的，从而该解决方案能够具有更灵活的综合过程。

在应用若干个估计策略并且对结果进行加权以导出用于所得到的与来自模拟形式的 CDP 的质量相关的判决的概率时，上述方法的组合是可能的。

关于打印-扫描环境和可能影响 CDP 质量的属性（见上）的信息可以按经编码的、机器可读的方式被存储在模拟形式上。或者，它可以被编码在 CDP 中。在这种情况下，将这样的关于打印机的信息结合到 CDP 中的一种方式是让打印机（例如桌面打印机）驱动器自动将唯一标识打印机的代码（“打印机标识符”）添加到被打印的 CDP 中。在扫描和验证一方，CDP 检测器读取打印机标识符并接收来自扫描仪的信息（或者是从扫描仪本身、本地或远程数据库获得的，或者是由用户输入的）。然后 CDP 检测器可以基于打印机标识符和关于扫描仪的信息从数据库（本地的或远程的）获得与打印-扫描环境相对应的校准数据以便进行判决。从而，可以对所有类型的打印机和扫描仪自动进行校准过程。

最后，在若干种情况下，可以进行模拟和数字表示形式之间的相关。以这种方式，完全避开了变换模拟或数字 CDP 的步骤。一个可能的相关函数是：

$$Dc = \sum_j \sum_i ((x(i,j) - \text{mean}_x) * (y(i,j) - \text{mean}_y)) \sqrt{\text{var}_x, \text{var}_y}$$

其中：

- mean_x 是 x 信号的平均值
- mean_y 是 y 信号的平均值
- var_x 是 x 信号的方差
- var_y 是 y 信号的方差

先前给出的技术可以单独应用到 VAP 的小部分，然后可以被聚集起来，以输出“全局得分”。例如，VAP 可以被划分成小块（例如 10x10 像素），并且相似性或相异性度量被应用到该块，以计算块的“块得分”。然后，各个块得分可以被聚集在一起，以用多种方式输出 VAP 的整体得分。一种明显的聚集是对所有块得分的简单平均，但是其他度量也可能是有利的；例如，可以按从最高值到最低值的方式挑选出块得分，例如，只有最佳的 25% 块得分被用于计算作为全局得分的平均值；或者全局得分被设置为第 25 个百分点处的块得分。

选择进入计算的块得分的有利之处在于可以从计算中排除 VAP 中更受模拟转换影响的区域。希望不会影响其得分的对 VAP 的更改为：磨

损，例如笔迹和纸张折叠；影响 VAP 的某些区域的聚焦不良的图像捕捉。如果不从全局得分的计算中排除这些更改 – 或者如果它们的影响不被最小化，则原始 VAP 可能具有太低的全局得分以至于不能被识别为原始的。

这里论述的技术还增大了对在打印和扫描期间通常发生的不那么明显、察觉不到的更改的鲁棒性。

在与来自模拟形式的 CDP 相比较之前利用关于打印-扫描过程的信息来修改来自原始数字表示形式的 CDP 的方法

一般，来自模拟形式的 CDP 被与来自原始数字表示形式的 CDP 相比较。但是，来自原始数字表示形式的 CDP 没有考虑到打印和扫描对来自模拟形式的 CDP 的作用。这些作用一般可以被描述为应用到来自原始数字表示形式的 CDP 的低通或带通滤波器。将来自模拟形式的 CDP 与已经被修改以模拟打印和扫描作用的来自原始数字表示形式的 CDP 相比较可以允许对来自模拟形式的 CDP 的质量进行更准确的测量。例如，正如一般观察到的，来自原始数字表示形式的 CDP 中具有低亮度的像素一般在来自模拟形式的 CDP 中显现为具有低亮度的像素；但是，如果在来自原始数字表示形式的 CDP 中来自模拟形式的 CDP 中的像素被亮像素所围绕，则其亮度一般会高于被暗像素围绕的情况。修改来自数字表示形式的 CDP 以模拟打印-扫描过程可以考虑来自模拟形式的 CDP 中的像素的邻居对该像素亮度的影响。

有若干种估计打印-扫描过程的方式。它们一般可以被划分成三个类别：

1. 可以通过查看大量来自模拟形式的 CDP 中的每个像素的平均亮度值来模拟打印-扫描过程。

2. 可以将打印-扫描过程估计为具有特定频率响应的滤波器。然后将该滤波器应用到来自数字表示形式的 CDP，并且在检测过程中使用经修改的来自数字表示形式的 CDP。

3. 打印机、纸张和扫描仪的物理属性有时可能是已知的，并且可用于

估计来自模拟形式的 CDP 的像素值。例如，热打印机在打印每个像素或点之后一般具有余热，并且这个余热可能对于之后的像素有影响，所述之后的像素可能被过热地打印，从而导致比起所需的像素来较暗。可以考虑这些物理作用，以及许多其他与特定打印和扫描设备和纸张的吸墨属性相关的作用。

获得由于打印和扫描导致的变换的参数

在下文中，公开了用于获得对由来自模拟形式的 CDP 中的打印和扫描导致的变换进行补偿所需的参数的多种不同技术。这些技术一般分成两个不同的群组：用于将参数信息包括在 CDP 本身或模拟形式中的技术以及用于训练 CDP 检测器的技术。

从模拟形式本身获得参数

从模拟形式的 CDP 获得参数

可以获得分析来自模拟形式的 CDP 所需的参数的一个地方是 CDP 本身。如前所述，CDP 可能携带消息，并且参数可以被包括 CDP 中携带的消息中。例如，就像直方图转换函数规格一样，一个质量阈值可以被存储在 CDP 中，在该质量阈值之上或之下来自模拟形式的 CDP 可以被判断为来自原始模拟形式或来自非原始模拟形式。

通用文档认证（始终真实）：每个打印机将新的 CDP 打印在每个被打印的模拟形式上。

打印机提供的 CDP 可以默认地打印在打印机上正打印的每个模拟形式上。如果打印机上打印的模拟形式是从具有 CDP 的原始数字表示形式制成的，则模拟形式将会具有两个 CDP：一个由打印机提供，一个来自原始数字表示形式。由打印机提供的 CDP 当然记录了由于将模拟形式打印在打印机上而导致的变换，并且这些变换可以用于确定模拟形式的来自原始数字表示形式的 CDP 是否指示模拟形式是打印-扫描-打印过程的结果，因此不是原始模拟形式。

关于打印过程的信息也可以被存储在远程数据库中，并且 CDP 或文档可以包含允许检测设备访问远程数据库中的该信息的标识。

CDP 和其他安全性特征

VAP 可以用“不可见墨水”打印在文档上，即，对于人类来说通常不可见但是具有容易检测的物理特性的薄的透明涂层，所述物理特性例如是导电性、电阻抗、电容或电致发光性。此外，可以用在安全文档部门广泛使用的专业墨水来打印 VAP。这些专业墨水包括从使用多种颜色到使用高强度紫外光来创建在可见或紫外光之下发荧光的图样中的任何一种。

物理安全性特征，例如纤维或墨水，可以携带用于 CDP 的密钥。CDP 也可以与诸如智能卡、2D 条形码、磁卡之类的安全性存储设备将结合。秘密密钥和依赖于打印机属性的校准数据可以被存储在这些存储设备中。

CDP 与物理、化学或光学防拷贝特征（例如出现在非原始模拟形式上的“VOID”缩放图（pantograph））也是互补的。

校准数据的自动选择

与打印机或扫描仪驱动器一样，CDP 读取器（验证器）可以手动地或自动地选择适当的校准数据。例如，当 CDP 被打印时，唯一地标识打印机或打印机的类别（打印机型号）代码被嵌入到 CDP 中，存储在数据库中或打印 CDP 的文档上。在验证阶段，读取器首先检测代码，然后选择适当的校准数据以用于验证。可以按类似的方式自动选择依赖于扫描仪的参数，例如验证的阈值。

自动校准 – 训练阶段

为了优化由打印机/扫描仪的固定设置限定的特定应用的 CDP 的检测，需要评估该应用的打印输出的扫描的各种参数。例如，上述直方图转换函数依赖于打印机和扫描仪设置，并且检测器进行判决所使用的质量或相似性阈值也依赖于扫描仪设置。但是，由于打印-扫描过程本质上是有噪声的，并且对于不同打印和不同扫描会在统计边界内变化，因此有必要从

统计上估计参数值。但是，打印和扫描三十个或更多个 CDP 以估计参数是困难、烦冗且易出错的。因此，使设置每个应用所需的这一过程自动化以允许任何人创建利用 CDP 检测拷贝的应用，很明显是有利的。

使从统计上估计参数的过程自动化的一种方式创建所谓的校准图像，该图像是包含从相同的数字表示形式制成的来自模拟形式的 CDP 或具有其中变化已知的特定密钥或有效载荷的 CDP 的若干个拷贝的数字图像。一个这种图像 1501 在图 15 中示出。该图像随后被以应用的打印机和扫描仪设置打印和扫描。然后，打印的校准图像被用 CDP 校准软件扫描和处理。如果校准成功，则校准软件输出应用的参数和检测器的判决阈值。判决阈值一般是通过以下方式计算的：首先测量质量指标的均值和方差，然后将统计模型拟合到这些数据以确定在 10000 个实例中不会导致例如 1 个差错的阈值（假定统计分布成立）。用户可给出假警报（将原本检测为拷贝）的概率的上界作为参数，其将被用于计算中以确定阈值。

每次新检测的迭代学习可以被结合到检测器中。在一个实施例中，可以要求用户使用具有五个不同的原始模拟表示形式的检测器。原始模拟表示形式的各种有效的统计数据 and 参数（例如判决阈值）被计算。在操作模式中，该判决阈值和其他参数和统计数据可以被用于判决，并被迭代地调节。

利用来自模拟形式的 CDP 本身来粗略确定模拟形式是否是原始模拟形式

为了测量模拟形式和原始 CDP 之间的质量差异，首先必须将通过扫描来自模拟形式的 CDP 制成的 CDP 的数字表示形式恢复到允许将来自模拟形式的 CDP 的数字表示形式与原始数字 CDP 相比较的形式。该过程通常成本较高，并且有时由于模拟形式的损坏而非常困难。在某些情况下，可能只需要对模拟形式是原始的还是非原始的模拟形式进行粗略的估计，或者就简单地是在没有成本高昂的相关（恢复和比较）过程的情况下不可能作出准确估计。其原因可能是计算成本太高，CDP 已经被损坏、缺少密钥或其他关键参数，或者检测软件由于安全性原因而不可用，等等。粗略估

计可以通过直接测量通过扫描来自模拟形式的 CDP 制成的 CDP 的数字表示形式的全局属性来进行：这些全局属性中的某些是直方图分布、平均亮度、两个连续像素之间的平均变化程度、二元 CDP 的黑像素和白像素的密度和分布，等等。由于这些全局属性一般在旋转、缩放和平移时是不变的，因此不需要恢复来自模拟形式的 CDP 的数字表示形式以测量这种全局属性。从而，对这些属性的测量既不需要原始数字 CDP，也不需要用来生成它的密钥。粗略估计是通过比较这些全局属性和依赖于 CDP 的随机性和高平均信息量的某些预定属性之间的相似性来进行的。

虽然判决模拟形式是原始模拟形式还是非原始模拟形式时的可靠程度不会像判决仅仅基于测量这些全局属性时那么高，但是大多数利用低端数字成像设备制作的非原始模拟形式是能够被检测到的。对来自模拟形式的 CDP 执行全局检查的检测器还可用于筛选出可疑模拟形式，并使其经历利用能够访问恢复来自模拟形式的 CDP 所需的信息的检测器进行的进一步调查。

作为前述技术的示例的邮资计算器应用

CDP 可以被自动插入到邮戳的数字图像中。然后图像一般被打印在粘到信封上的粘贴物上，或者直接打印在信封上。来自邮戳的 CDP 随后可被用于自动检测邮戳是原始模拟形式还是非原始模拟形式。利用来自邮戳的 CDP 测量的质量指标可以与邮戳的其他特征一起，例如确定产生邮戳的打印机、分析字母的字体和/或读取打印的数字水印，被结合到用于自动地或手动地确定邮戳是否是原始模拟形式的全局得分中。

打印的邮戳上的 CDP 可以用于法庭验证，例如当可疑邮戳被拿到配备有平头扫描仪和检测软件的站时。当高速扫描仪（例如 WFOV）对每个邮戳进行图像捕捉时，它还可以用于自动验证。

在这种应用中，用于 CDP 的密钥可以是固定的或可变的。如果它是可变的，则密钥可以（部分地）从邮戳中包含的其他信息导出，所述其他信息例如是发送者名称。

用于将信息插入到 VAP 中的技术

如前所述, VAP 可以携带信息。当 VAP 被创建时, 消息被结合到 VAP 中。由于 VAP 是以消息创建的, 因此每条不同的消息导致不同 VAP 的生成。将消息结合到 VAP 中的一种方式是利用秘密密钥和消息作为用来生成 VAP 的伪随机数生成器的输入。一般希望以保持 VAP 的直方图或平均信息量的方式将消息结合到 VAP 中。所有前述用于将消息结合到 VAP 中的方案都具有与不包含信息的 VAP 类似的直方图和平均信息量。但是即使保持 VAP 的直方图或平均信息量能够有某些优点, 例如 VAP 辨别拷贝和原本的能力不应当受到影响, 但是保持平均信息量和直方图并不是必要条件。换言之, 在不保持 VAP 的平均信息量或直方图的情况下仍有可能创建包含信息的 VAP 并利用它们来辨别拷贝和原本。

在一个实施例中, VAP 被划分成不同的不交迭的区域, 例如划分成相邻的固定大小的块。大小是任意的, 例如 1x1 像素、4x4 像素或 8x4 像素都可以。如果块大小太大, 则消息大小将会受限; 如果太小, 则消息将会难以读取。例如, 如果每个块携带消息的单个比特, “1” 或 “0”, 则 VAP 中的每个像素或者属于保存 “1” 的像素的类别, 或者属于保存 “0” 的像素的类别。用于伪随机数生成器中的像素所属的类别 “c”、其位置 “p” 和秘密密钥 “key” 被用于函数 G 中, 以产生像素值 x, 即:

$$x=G(\text{key},p,c)$$

对于二元像素值, x 是 0 或 1, 对于灰度级像素值, x 是 0 到 255 之间的整数。x 的所有值都等概率的。

可以将函数 G 分解成两个函数 G1 和 H, 其中 G1 是以密钥和像素位置作为输入的伪随机数生成器, H 是根据像素所属的类别调节像素值的函数:

$$x=H(G1(\text{key},p),c)$$

在一个可能的实现方式中, H 是这样的: 当块携带比特 “0” (或者换言之其类别为 “0”) 时, H 保持像素值不变:

$$H(G1(\text{key},p), '0')=G1(\text{key},p)。$$

但是当块携带 “1” 时, H “颠倒” 像素值, 即:

如果像素值是二元的，则

$$H(G1(key,p), '1') = 1 - G1(key,p)$$

如果像素值是灰度级的，则

$$H(G1(key,p), '1') = 255 - G1(key,p)$$

描述块如何能够携带消息元素的另一种方式是，通过限定块的两个类别：B1 和 B0，一个块中可以携带一个比特。B1 是携带比特“1”的块类别，B0 是携带比特“0”的另一块类别。B1 和 B0 优选地是以这样的方式定义的，即类别为 B1 的任何块（b1）与类别为 B0 的任何块（b0）具有最小的相关性。创建 b0 或 b1 的像素值的一种简单的方式是将比特值作为除密钥和像素位置之外的随机数生成器种子。为了减少 b0 和 b1 之间的潜在“交迭”，尤其是对于具有少量像素的块，可以按 b0 的值与 b1 的值很不相同的方式来控制随机数生成器。在检测期间为了读取比特值，来自模拟形式的块被与 b1 和 b0 相比较。如果与 b1 的相关性好于与 b0 的，则比特值“1”被读取。否则比特值“0”被读取。

有许多方法可以用来将来自模拟形式的块与 b0 和 b1 相比较；但一般来说，相关性或距离度 m 量将会提供满意的结果。更一般地说，如果块类别数从 2 增大到 N，则块可以对 N 个值中的任何一个进行编码。注意在 N 个块类别的情况下，块可以包含多达 $\log_2(N)$ 个比特。例如，如果 $N=4$ ，依赖于块属于四个类别中的哪一个，具有“1”值的块可以对 0 到 3 之间的值进行编码，因此包含 2 个比特。为了使该技术对于 N 个值也能起作用，在代表值的类别的块和不代表值的类别的块之间需要最小的相关性。

“块”的定义可以延伸到代表从 1x1 像素到像素的任意不连通集合中的任何一种。

图 17 示出插入了消息的一组像素的变换。具有未被修改的像素值的块在 1701 处示出。在 1702 处是在块被修改以携带具有值“1”的比特之后 1701 处所示的像素块。

易于验证，以上述方式插入了信息的 VAP 的像素值的直方图或频率分布保持不变。图 12 在 1201 处示出了利用密钥“test”生成的 VAP，在 1203 处示出了利用相同的密钥生成并添加了信息（整数值 123456789）的

VAP, 在 1205 处示出了差异图像。差异图像的黑区域（实际上是 4x4 像素的块）对应于未被信息插入所修改（颠倒）的像素值：它们对应于嵌入“0”的块。很明显，类似噪声的经修改的区域对应于嵌入“1”的区域。

执行检测的一种可能的方式包括在第一步骤中生成不包含信息的数字 VAP。然后经扫描、恢复和直方图转换的 VAP 的每个块被与（1）原始数字 VAP 的相应块和（2）具有颠倒的像素值的相同块相比较。可以使用不同的比较函数：欧几里得距离、绝对距离等等。然后，如果最接近的块是经颠倒的那个，则该块的比特值被假定为“1”，否则被假定为“0”。

例如，图 17 中的 1703 示出从模拟形式扫描、恢复和转换后的像素块。当该块与来自原始数字表示形式的 VAP 的相应块和具有颠倒的像素值的相同块相比较时，绝对距离为：

对于“0”： $(|243-44|+|228-36|+...+|104-172|)/16=132.81$

对于“1”： $(|12-44|+|27-26|+...+|151-172|)/16=22.93$

在这种情况下，检测到的比特为“1”。

如果某个比特在 VAP 中的不同位置被嵌入若干次，则可以跟踪每个位置处的“0”和“1”的距离，以便每个块对最终关于比特值的判决的贡献被加权。这样，其中对于“0”发现距离为 55.32 而对于“1”发现距离为 51.34 的块的贡献将会小于其计算如上所示的块，其中有利于“1”的评剧更强（距离 22.93 对 132.81）。

由于消息的存在影响 VAP 的外观，因此，如果消息的结合对 VAP 的影响未被考虑到，则包含消息的扫描的 VAP 就不能与来自原始数字表示形式的 VAP 相比较以确定扫描的 VAP 是来自原始模拟拷贝还是来自非原始模块拷贝。这可以在一旦消息已知时通过以下方式来完成：重新创建具有消息的来自原始数字表示形式的 VAP，然后将重新创建的具有消息的 VAP 与扫描的具有消息的 VAP 相比较。在某些情况下，消息不能被读取，这例如是因为打印的质量太低。但仍希望确定 VAP 是否是经损毁的原本，或拷贝，或者甚至是看起来像 VAP 但却是以不同方式（例如利用不同的密钥）生成的图样。为了允许进行该确定，可以预留 VAP 中的一部分，这一部分的块不包含消息元素。块的该部分可以针对给定秘密密钥

确定，并且该部分可以用于进行确定，即使消息不能被读取也是如此。

在每个块中检测到的个体比特值可以用于以鲁棒且安全的方式对特定消息进行编码。例如，为了将由标识数字“123456789”构成的消息嵌入到 VAP 中，可以使用下述过程：

- 以 4 字节或 32 比特表示消息，这允许了表示 2^{32} 个整数值中的任何一个。
- 向 32 比特消息添加差错校正/差错检测码；如果使用(8,28)BCH 码，则带有差错校正码的消息由 $28 \times 4 = 112$ 比特表示
- 计算 VAP 中的可用块的数目；对于划分成 4 块的 100×100 像素图像，这总共是 625 块。112 比特中的每一个可以在 VAP 中的 5 个位置处插入（ $625 - 112 \times 5 = 65$ ，剩下的块可以不用，或用于第 6 次嵌入某些比特）。
- 利用秘密密钥对 112 比特流加扰，并将其插入到 VAP 的伪随机选择的块中，其中利用上述方式之一来插入比特 1 或 0。

为了从 VAP 的模拟表示形式检测消息，一般可以按相反的顺序来执行步骤。注意对于非原始 VAP 频率是相同的。

- 确定如果嵌入了 0 或嵌入了 1 则每个块中 VAP 的值应当是什么；
- 对于模拟 VAP 中的每个块，测量两个得分，一个对应于块包含“1”的可能性，另一个对应于它包含“0”的可能性；
- 由于消息的每个比特一般被表示若干次（在这种情况下 112 个比特被表示 5 次），因此累积来自嵌入比特的不同位置的个体得分（这些位置可以用秘密密钥来得知）。基于累积得分选择最可能的比特。
- 向 112 个比特应用差错校正，以得到 32 比特消息
- 将 32 比特消息转换为整数，在这种情况下为“123456789”。

分布式 VAP

对于某些文档，VAP 的可视方面可能与审美要求是不兼容的。例如，钞票的审美方面是非常重要的，并且一般这些文档的安全性特征必须是不明显的，或者是不会产生干扰的，或者自然地融合到文档设计中。向这种文档应用 VAP 的一种方式是将它们分布在文档上。图 14 示出了具有分布

式 VAP 1403 的支票 1401。分布式 VAP 由多个 VAP 单元 1405 构成。在支票 1401 中，每个 VAP 单元 1405(i)的大小为 10x10 像素，并且 VAP 单元 1405 在分布式 VAP 1403 中按 100 像素的间隔规则分布。当放在一起时，VAP 单元 1405 中的像素等同于大小为 240x100 像素的非分布式 VAP。支票 1401 的 VAP 单元是很明显的，但是通过利用纹理更重的 VAP 单元、通过利用较小的 VAP 单元和/或通过使 VAP 单元的位置随机化，可以使它们不那么明显。当创建 VAP 单元时，可以向单元的像素赋予某个范围的值，该范围与文档在单元在文档中的位置处提供的背景相混和。还可以调节文档的给定区域中的 VAP 单元的密度，以使 VAP 单元更不易察觉。可以用秘密密钥来确定 VAP 单元的位置。

具有像素大小或点大小的 VAP 单元的分布式 VAP

分布式 VAP 可以由像素大小或点大小的 VAP 单元构成。在这种分布式 VAP 中，像素大小或点大小的 VAP 单元（在下文中都称为“点”）分散在模拟形式上，并且每个点具有随机的像素亮度值。点可以按规则图样分布，或者可以具有随机或伪随机的分布。分布可以用秘密密钥来确定。像其他 VAP 那样，点是在模拟形式被打印时打印的。图 16 示出了具有随机分布的点 1602 的模拟形式 1601。伪随机点和点值的图样可以是固定大小的，例如 2x2 英寸，并且图样可以平铺在整个模拟形式上。显然，多个 CDP 数字像素必须在应用中的使用的打印机上以相同的打印参数被打印，而且也必须用应用的扫描仪以相同的参数扫描。如果没观察到这一点，或者使用了错误的阈值和参数，则可能导致不那么可靠的检测，或者完全错误的结果。

在检测时，可以应用点检测算法来检测模拟形式的数字化图像中的所有点或大部分点。用于定位点的技术在 8/24/04 授权的 Zhao 等人的标题为 *“Apparatus and methods for improving detection of watermarks in content that has undergone a lossy transformation”* 的美国专利 6,782,2116 中描述，这里通过引用将其结合进来用于所有目的。一旦检测到一组点，就应用搜索算法来找出具有特定属性的点群组。一个这种属性可以是点的像素的值与文

档在点位置处提供的背景之间的关系。一旦找出这种点的一个群组，点图样的配准就已知了，来自模拟形式的点的图样就可以与来自原始数字表示形式的点的图样相比较，并且模拟形式的质量指标就可被确定。

在包含文本和图形的区域上分布分布式 VAP

通过将分布式 VAP 的像素分布在包含文本或图形的区域上，可以使之更不明显，如图 16 中的 1603 处所示。这种 VAP 可以大体上针对分布式 VAP 按所述方式来处理。分布式 VAP 的单元也可以形成文档中的特定可视图样，例如可见水印。

在包含其他安全性特征的区域上分布分布式 VAP

分布式 VAP 的像素可以用于形成安全性特征，例如缩放图、活动图、全息图或微缩文本（微缩打印）。

利用非灰度级打印机打印 VAP

某些类型的打印机只能打印纯黑像素，而不能够打印具有 0 到 255 之间的任何可能的像素亮度值的灰度级 VAP。对于这类打印机，用于生成 VAP 像素值的伪随机数生成器可以被设置成使得像素值或者是“0”（黑）或者是“1”（白）。上述算法均可应用于二元 VAP。

应当注意，某些打印机，比如喷墨或激光打印机，实际上通过采用数字中间色方法产生了某个范围的灰色调，即，通过以高分辨率（例如 1200dpi）打印微小的二元（黑？）点获得较低分辨率（例如 300ppi）的灰度级图像。对于这些打印机，图像的二元表示形式是由打印机从输入的灰度级表示形式产生的，最终它是被打印的二元黑白图像。不是生成和打印必须经历打印机的变换的灰度级 VAP，可以产生由打印机按原样打印（没有数字中间色过程）的较高分辨率（例如 1200dpi）的二元 VAP。

为了降低较高打印分辨率下黑像素之间潜在的干扰（主要是由墨的融化或纸张对墨的吸收造成的），可以在创建 VAP 时控制黑白比率。对于其中黑像素和白像素的数目几乎相等的正常二元 VAP，比率应当约为

1.0。降低的比率导致具有较少像素的 VAP。出于上述原因，对于某些应用这可能更为适当。

要打印二元 VAP，重要的是要使用适当的打印分辨率，即每英寸像素（PPI）。如果 PPI 被设计得太高，则打印机和/或打印介质（即纸张）可能不能够提供这种分辨率，从而原始模拟形式的质量可能低于预期。另一方面，如果 PPI 被设置得太低，则在模拟形式上打印的每个像素可能很容易被扫描仪拾取，从而质量恶化量可能太小，以至于不能辨别拷贝和原本。选择适当的 PPI，以便以最佳的质量打印 CDP。因此 PPI 常常依赖于打印机的分辨率和用于特定应用的打印介质的质量。

最后，存在其他能够打印二元 VAP 的二元打印过程：例如，金属、玻璃塑料或其他材料的激光雕刻、某些全息图，等等。

将 VAP 用于可视加密

VAP 的一个有趣的属性是，除了它们自动辨别拷贝和原本的用途之外，它们还可以用于被称为可视加密的过程中的手动认证，其中可视加密首先在 *Moni Naor and Adi Shamir, Visual Cryptograph, Eurocrypt 94* 中描述。

为了能够利用 VAP 进行可视加密，按通常的方式创建和打印 VAP。让我们把打印的 VAP 称为 VAP0。然后，需要判定在合法的 VAP 中希望看到哪个可视消息（或图像）。一般来说，可视消息是数字黑/白图像，它与数字 VAP 具有相同的像素大小（可视消息总是能被调节到匹配数字 VAP 的大小）。让我们称该可视消息为 VM。然后，创建与 VAP0 和 VM 具有相同的像素大小的被称为 VAPk 的新 VAP，以使得对于每个像素 (x,y)：

$$\begin{aligned} \text{VAPk}(x,y) &= 1 - \text{VAP0}(x,y) && \text{如果 } \text{VM}(x,y) = 0 \\ \text{VAPk}(x,y) &= \text{VAP0}(x,y) && \text{如果 } \text{VM}(x,y) = 1 \end{aligned}$$

VPKk 是可视密钥：如果它是在透明片上打印的（值为“1”的像素将会使光通过），则当它被准确地放置在打印的 VAP 之上时，它会揭露消息。实际上，在这种情况下，观察者所感知到的亮度值（我们将其称为

$I(x,y)$ 将会是打印的 VAP 和 VAP 手动密钥中最暗的那个:

$$I(x,y)=\min(VAP(x,y),VAPk(x,y))$$

易于验证, 在消息的黑区域中 (其中 $VM(x,y)=0$), $I(x,y)$ 将始终为 0, 而在消息的白区域中, $I(x,y)$ 将具有与打印的 $VAP(x,y)$ 相同的值。对于具有等概率的值 0 和 1 的二元 VAP, 这意味着在消息的白区域中, $I(x,y)$ 的 50% 的像素将会具有明亮值。总的来说, 在观察者看来, 可视消息的对比度比其原始版本稍小, 但仍是可辨别的。

图 18 示出二元 VAP0 1801, 可视消息 “1234” 1805, 从二元 VAP 和可视消息创建的可视密钥 VAPk 1803。最后, 当 VAPk 1803 被覆盖在 VAP0 1801 上时, 可视消息被揭露的方式。

一般原理

引起 VAP 的上述改进的工作也阐明了 VAP 中体现的多个关于拷贝检测的一般原理。VAP 中体现的一个一般原理是由拷贝过程产生的假象可以用于确定数字表示形式已经经历了多少次拷贝过程。该原理的一个示例是利用由在制作非原始模拟形式时涉及的打印和扫描所产生的假象来检测模拟形式是否非原始。正如下文将更详细描述, 该原理的另一个示例是使用数字表示形式中不具有差错校正的一部分来确定整个数字表示形式已经经历了多少次拷贝过程。两个示例的根本之处都在于存在禁用差错校正的拷贝。在所关注的是涉及数字和模拟形式之间的变换的拷贝的情况下, 数字差错校正对于模拟域没有影响; 在数字到数字拷贝的情况下, 必须规定在数字表示形式的一部分中禁用数字差错校正。

VAP 中体现的另一个一般原理是用于拷贝检测的区域具有平均信息量较高的随机图样。该图样是由图样元素构成的, 这些图样元素可以采取任何形式, 但是一般是包含 VAP 的数字表示形式的素元。例如, 如果数字表示形式是图像, 则图样元素可以是像素, 如果它是视频, 则图样元素可以是用于代表视频图像的块。图样一般是用秘密密钥产生的。图样的随机性和平均信息量使得难以看到由拷贝产生的假象, 因此难以修改图样以隐藏由拷贝过程产生的假象。随机图样的属性被仔细地选择, 以考虑到由拷

贝过程产生的假象的特性。在 VAP 的情况下，例如，所选择的属性一方面从拷贝过程幸存，另一方面却受到它的显著影响。

用于涉及数字和模拟域之间的变换的拷贝的认证图样

在下文中，这种认证图样将被称为模拟-数字认证图样。

模拟-数字认证图样的一般原理

模拟-数字认证图样基于下面两个原理：

1) 非原始模拟形式始终经历模拟输出过程的模拟输出-数字记录，而原始模拟形式仅被输出。这导致了原始模拟形式和非原始模拟形式之间的可检测的质量差异。

2) 模拟-数字认证图样 (ADAP) 是以密钥生成的图形，它被设计为对由模拟和数字域之间的转换产生的变换具有最大程度的敏感性，并且被设计为以使得在来自模拟形式的 ADAP 的数字化表示形式上工作的自动检测器能够测量 ADAP 的质量恶化指标从而能够确定模拟形式是原始还是非原始模拟形式的方式示出这些变换。

容易看出，这两个原理可以应用到任何从原始数字表示形式产生的模拟形式。在所有情况下，模拟形式都是从数字表示形式制成的。任何这种模拟形式都可能包含伪随机有噪声信号，这是图像中 VAP 的功能等同物。ADAP 从而是不同模拟形式的 VAP 的等同物。类似地，CDP 的等同物可被称为拷贝检测信号 (CDS)。

与 VAP 一样，ADAP 所要求的是 ADAP 能够与模拟形式的其他内容相区分。例如，对于无线电信号，某些频带可以携带 ADAP。对于诸如磁或光带、盘或条纹之类的介质，带、盘或条纹的预定的区域可以包含 VAP，而对于多媒体模拟形式，ADAP 可以包含在多媒体模拟形式的一个轨道中。

ADAP 的示例性应用

利用 CDS 检测伪造的 RFID 信号

射频识别设备（RFID）被附着到任意个体物品，并且发射充当物品的唯一标识的唯一信号。虽然该技术刚刚起步，并且 RFID 的成本仍然较高，但是预期该技术的各种应用将会出现，尤其是在零售、发行和存储领域。

对于造假者，RFID 的广泛使用意味着伪造品必须有发射伪造的信号。RFID 信号的伪造可以通过捕捉原始 RFID 信号并创建再现捕捉的 RFID 信号的 RFID 设备来完成。

与用 VAP 来检测非原始模拟形式的情形的相似性是显著的：假定通过制作要伪造的 RFID 信号的原始模拟形式的数字化表示形式来捕捉要伪造的 RFID 信号，并且将数字化表示形式用于产生伪造（非原始）模拟形式，然后，RFID 信号的原始模拟形式在其真实性被分析之前经历一次模数变换；非原始模拟形式经历了两次模数变换，一次是在产生伪造模拟形式的过程期间，另一次是在其真实性被分析之前。额外的模数转换一般会导致 RFID 信号的质量或信息的额外损耗。因此，可以创建拷贝检测信号（CDS），该信号对于由模数转换产生的变换具有最大程度的敏感性，并且将被 RFID 发射。该信号将被 RFID 检测器分析，以按与分析 VAP 很相同的方式检测 RFID 设备是真实的还是伪造的。

利用 ADAP 大体上确定信号经历了多少次数模转换和模数转换

每次信号经历 A/D 或 D/A 转换时，转换设备将会引入“噪声”。一般来说，这样引入的噪声具有特定的“噪声”图样。通过在信号中包括 ADAP 以测量由转换引入的噪声，分析“噪声”的特性，并将 ADAP 与原始 ADAP 相比较，可以确定是否将包含 ADAP 的模拟形式或数字表示形式视为剽窃拷贝，还可以确定关于用来制作剽窃拷贝的机制的某些信息。此外，还可以通过测量质量恶化量了解拷贝的生成。

利用 CDS 堵塞数字表示形式的防拷贝的“模拟漏洞”

“模拟漏洞”是指打算用模拟装置重放的数字内容的防拷贝方案中的一个基本弱点。当信息被转换到模拟形式时，所有数字拷贝防护机制（数

字表示形式的防拷贝) 都丢失了, 对于产生的模拟信号没有限制, 并且内容能够在无限制的情况下被捕捉回数字形式。通过向要保护的数字表示形式添加 CDS, 设备 (例如 DVD 播放器或 DVD 记录器) 可以包含 CDS 检测器, 该 CDS 检测器自动读取 CDS 的质量, 并且如果 CDS 检测器指示数字表示形式是原始数字表示形式或非原始数字表示形式, 则许可或拒绝操作。CDS 可以按多种方式放置在多个位置。例如, 它可以在电影开始处以“著作权警告”帧的形式以可见方式插入, 或者可以以分布式 PAP 的形式不明显地放置在电影的特定位置。该原理还可用于检测在电影院中以便携式摄像机捕捉并被转换和压缩到数字文件的电影拷贝, 这是因为便携式摄像机记录也被称为“拷贝”过程。当尝试播放或拷贝用便携式摄像机剽窃的电影时, 包括 CDS 检测器的设备将会停止播放或拷贝。此外, 如果所有便携式摄像机都配备有 CDS 插入器, 则在记录期间这种 CDS 插入器将会自动将 CDS 插入到某些帧中, 同时这种 CDS 携带唯一标识便携式摄像机 (例如便携式摄像机 ID) 的消息。从而, 从用便携式摄像机剽窃的电影中, CDS 检测器不仅检测到它是拷贝, 而且还识别出剽窃的源。

利用认证图样在数字介质上实施“一次拷贝”规定

当原始信息以数字形式被记录到诸如计算机硬盘、数字带、高密盘或 DVD 盘之类的存储介质时, 额外的盘级别差错校正码 (例如 Reed Solomon 码) 一般被添加以校正记录或读取设备或存储介质的物理损毁 (例如 CD 或 DVD 的刮擦) 可能引入的差错。由于盘级别差错校正, 在校正上述差错后, 可从存储介质读出与原始信息完全相同的数字表示形式。也可以从存储介质制作原始信息的完全相同的拷贝。但是, 如果没有这种盘级别差错校正, 从其中存储了原始信息的存储介质制作的拷贝将会由于记录、读取和介质本身引入的差错而经历“恶化”。由于即使数字到数字拷贝也会导致拷贝相对于原本的恶化, 因此 CDS 可以用于实施许可用户进行单次数字拷贝的“一次拷贝”规定。可以利用密钥以伪随机方式生成 CDS, 这使得其内容不可预测。为了确保 CDS 将由于拷贝过程而质量降低, 它不包含差错校正方案。例如, 可以被插入到高密盘或 DVD 的

“用户数据”部分，插入/分布到视频或音频压缩格式（例如 MPEG-2 或 H.264）的“用户数据”字段，或者作为原始信息的一部分（例如电影开始处的视频帧）插入或分布。当制作原始数字表示形式的拷贝时，CDS 将被以质量降低的方式再现。如果制作拷贝的拷贝，则 CDS 将会经历两次质量降低，从而具有更低的质量。给定 CDS 的这个质量，将构造出这样的设备，其自动地读取 CDS 的质量，并且只在 CDS 指示数字表示形式是原始数字表示形式而本身不是非原始数字表示形式时才许可用户制作拷贝。这种拷贝检测信号的一个使用领域是只许可作为原始数字表示形式的 DVD 被拷贝的拷贝设备上的拷贝检测机中。

结论

前述“*具体实施方式*”向相关技术领域的技术人员公开了发明人的利用拷贝检测信号来主要确定数字表示形式是原本还是拷贝的技术，用于确定对象的模拟形式是原始模拟形式还是非原始模拟形式的技术，它们的经改进的用于定位和分析 VAP 的技术，以及它们的用于将消息包括在 VAP 中的新技术，并且还向相关技术领域的技术人员公开了发明人目前知道的用于实践这些技术的最佳模式。相关技术领域的技术人员立即可以明白，除了这里公开的那些实施例外，申请人的技术的许多其他实施例也是可能的。拷贝检测图样的基本技术可以和任何数字介质以及图样元素、尺寸、形状、位置一起使用，并且拷贝检测图样的图样以及用于分析拷贝检测信号的技术将由与拷贝检测图样一起使用的介质的性质以及拷贝检测图样的目的确定。VAP 或 CDS 如何携带额外的信息，该信息是什么，以及它如何用于分析过程中也将由应用 CDS 的介质的性质以及应用的目的来确定。一般来说，CDS 可以用于任何要检测由拷贝过程导致的变化的情形中。

由于以上所有原因，“*具体实施方式*”应当被视为从任何意义上来说都是示例性的而不是限制性的，并且这里公开的发明的范围不由“*具体实施方式*”确定，而是由以专利法所许可的完整范围解释的权利要求书来确定。

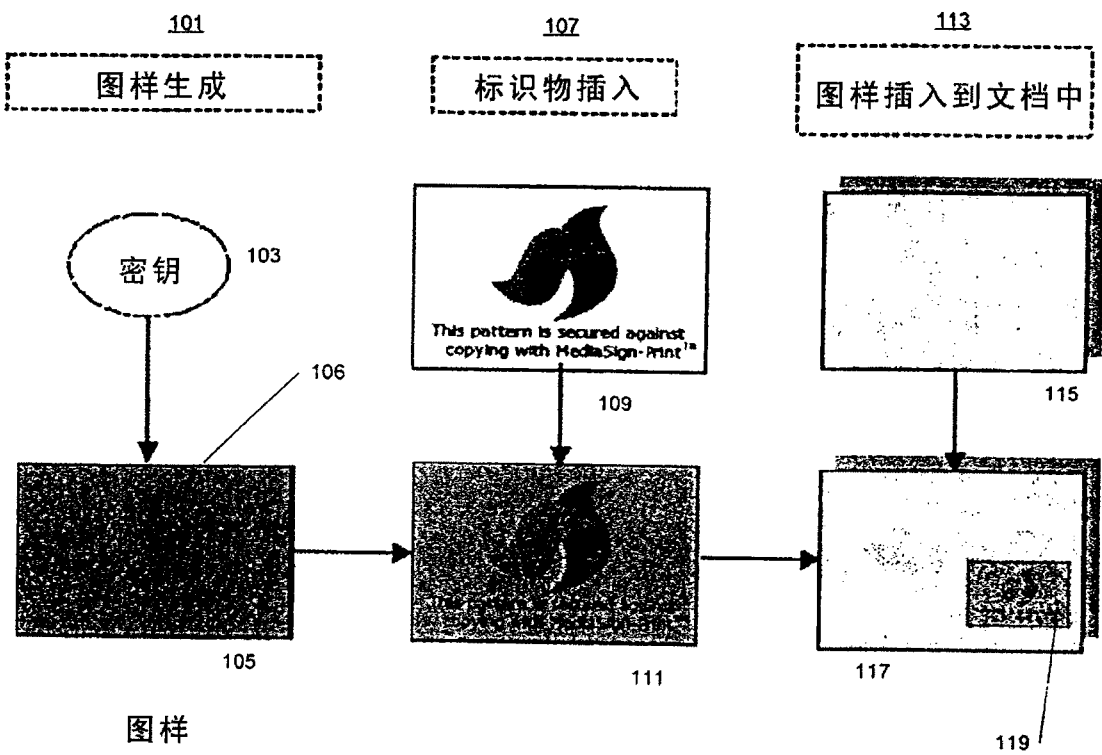


图1

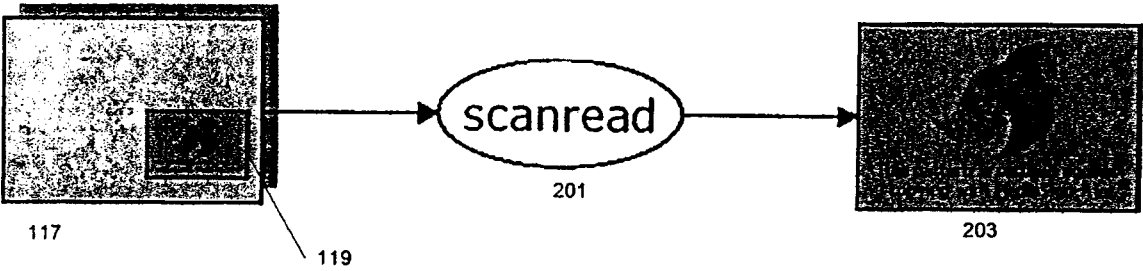


图2

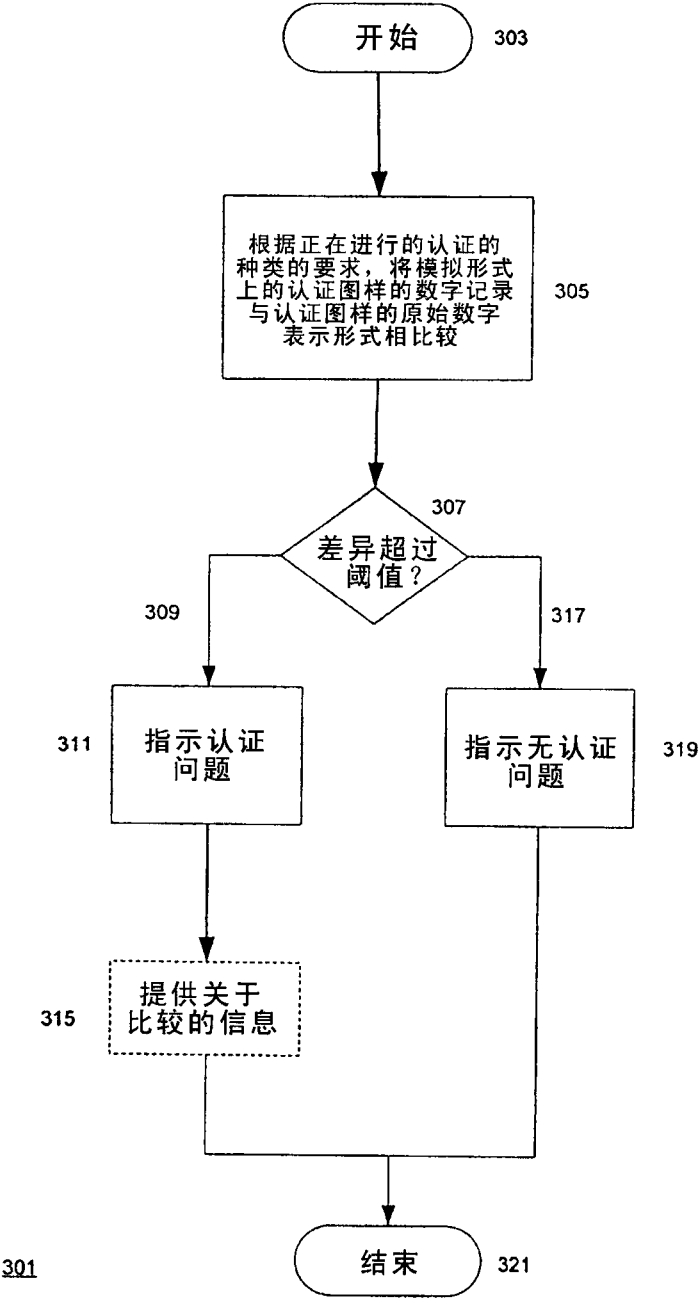
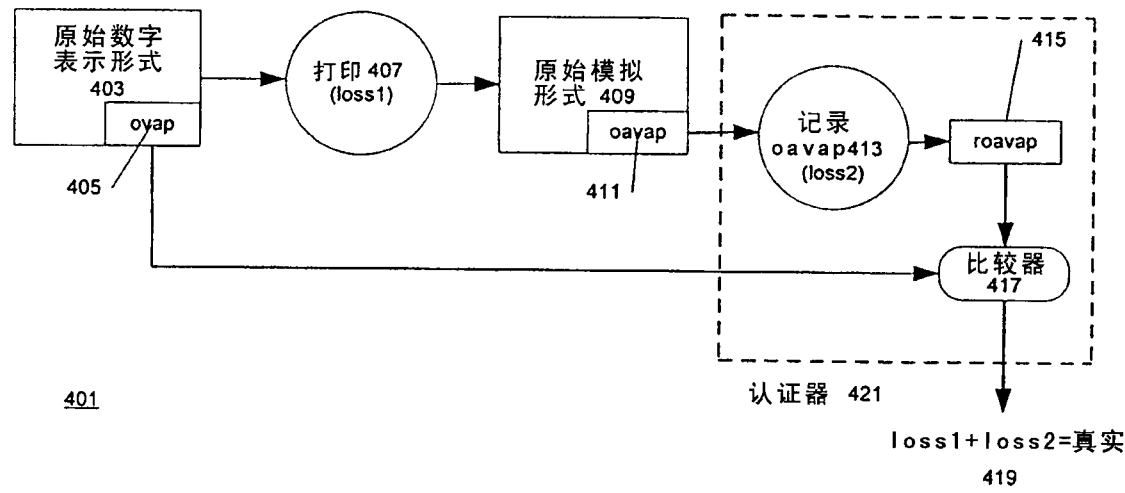


图3

场景1：原始模拟形式的打印和认证



场景2：非原始模拟形式的打印和认证

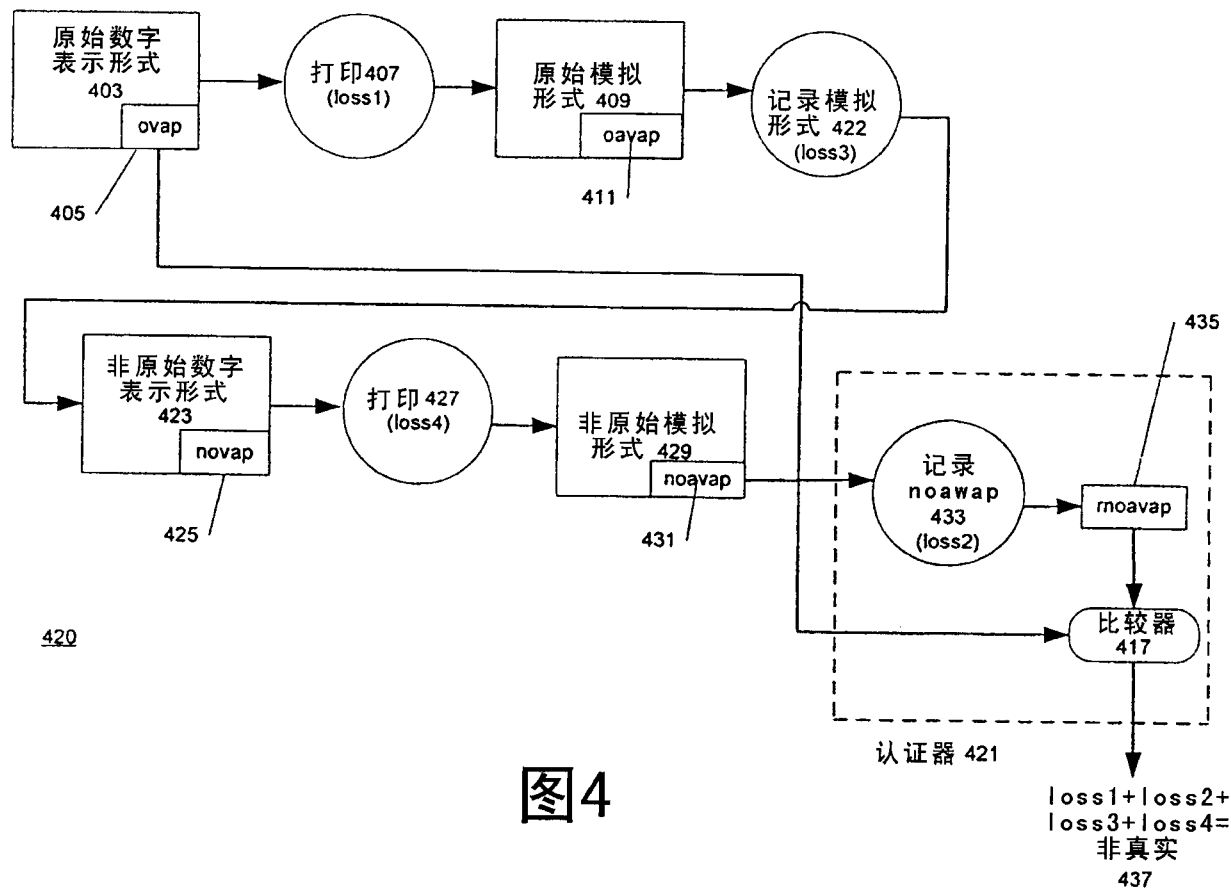
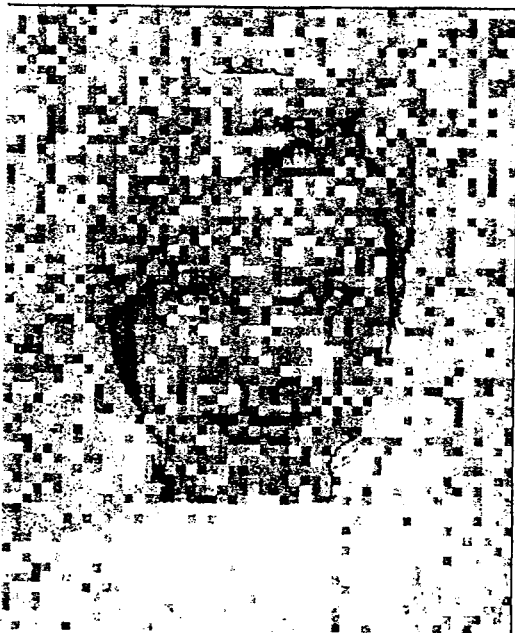


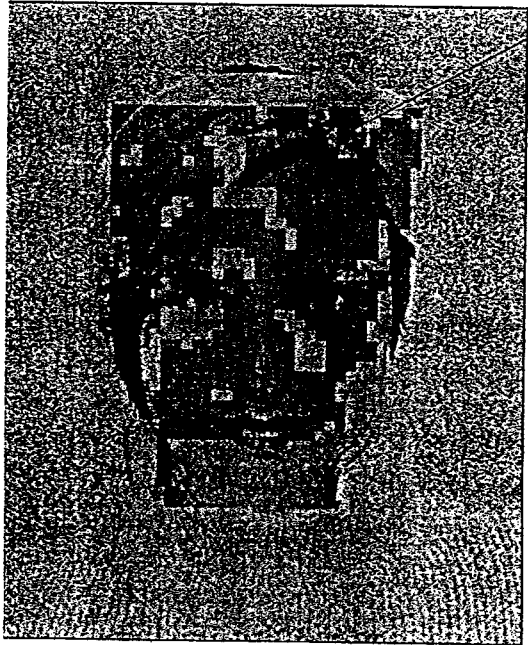
图4



501



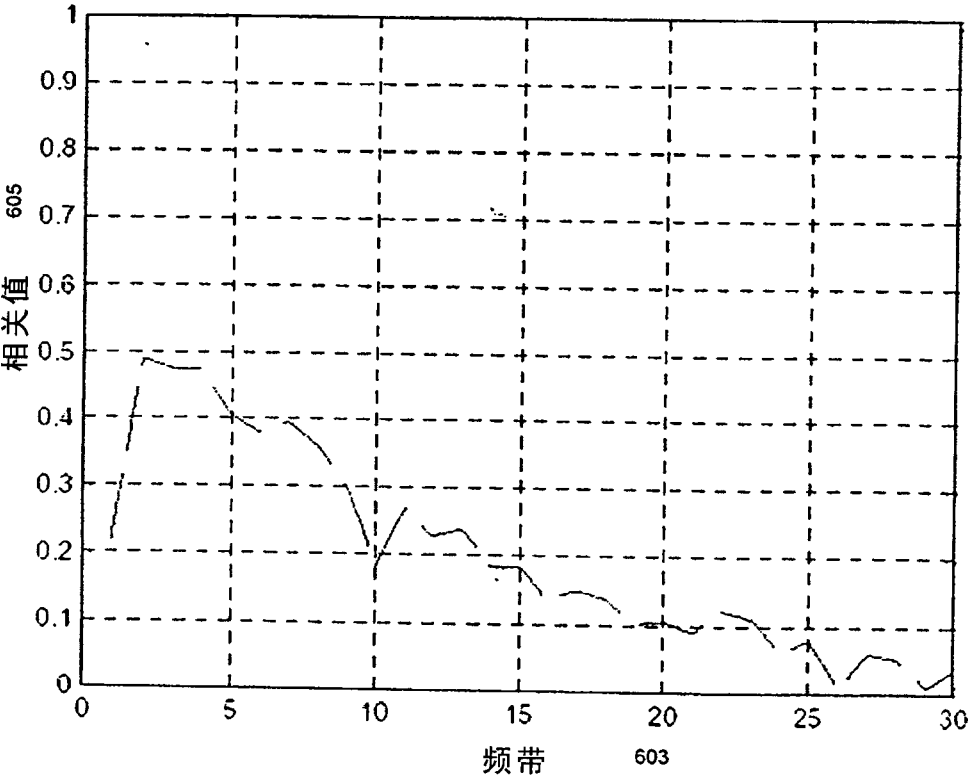
502



503

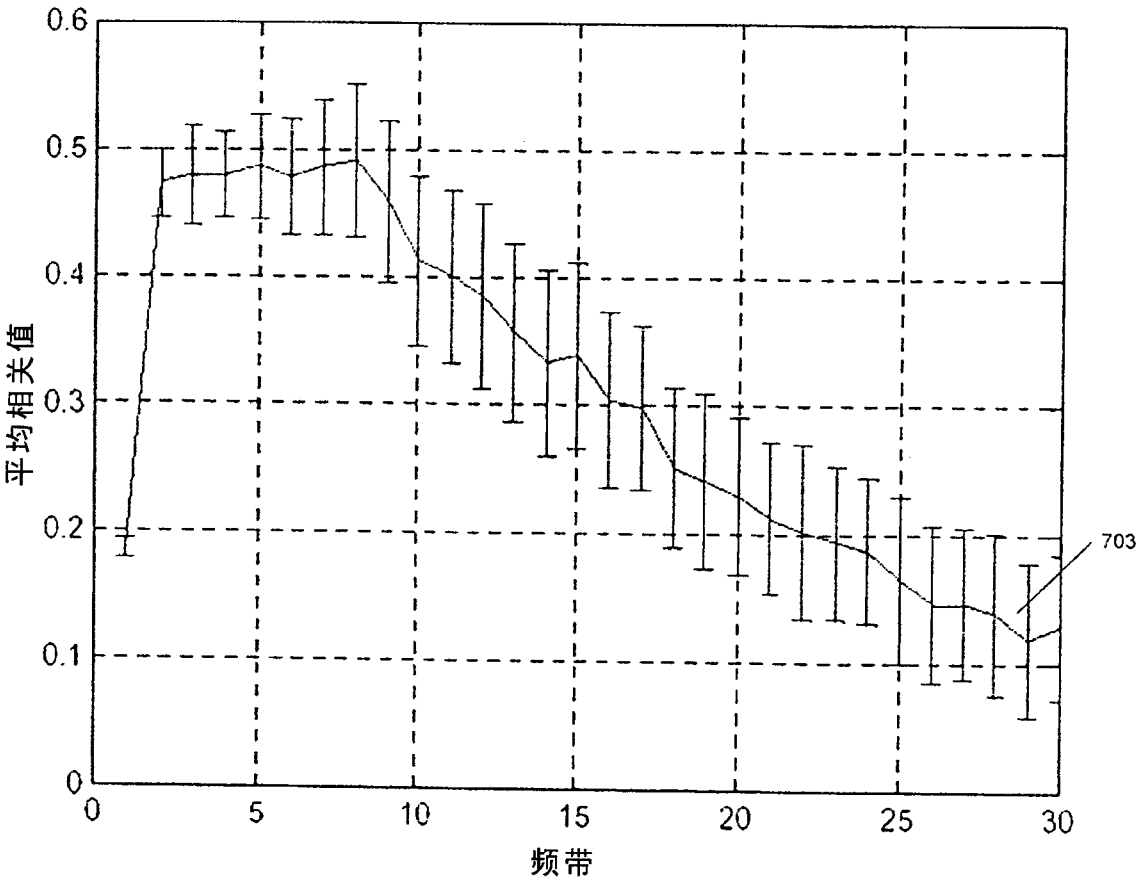
505

图5



601

图6



701

图7

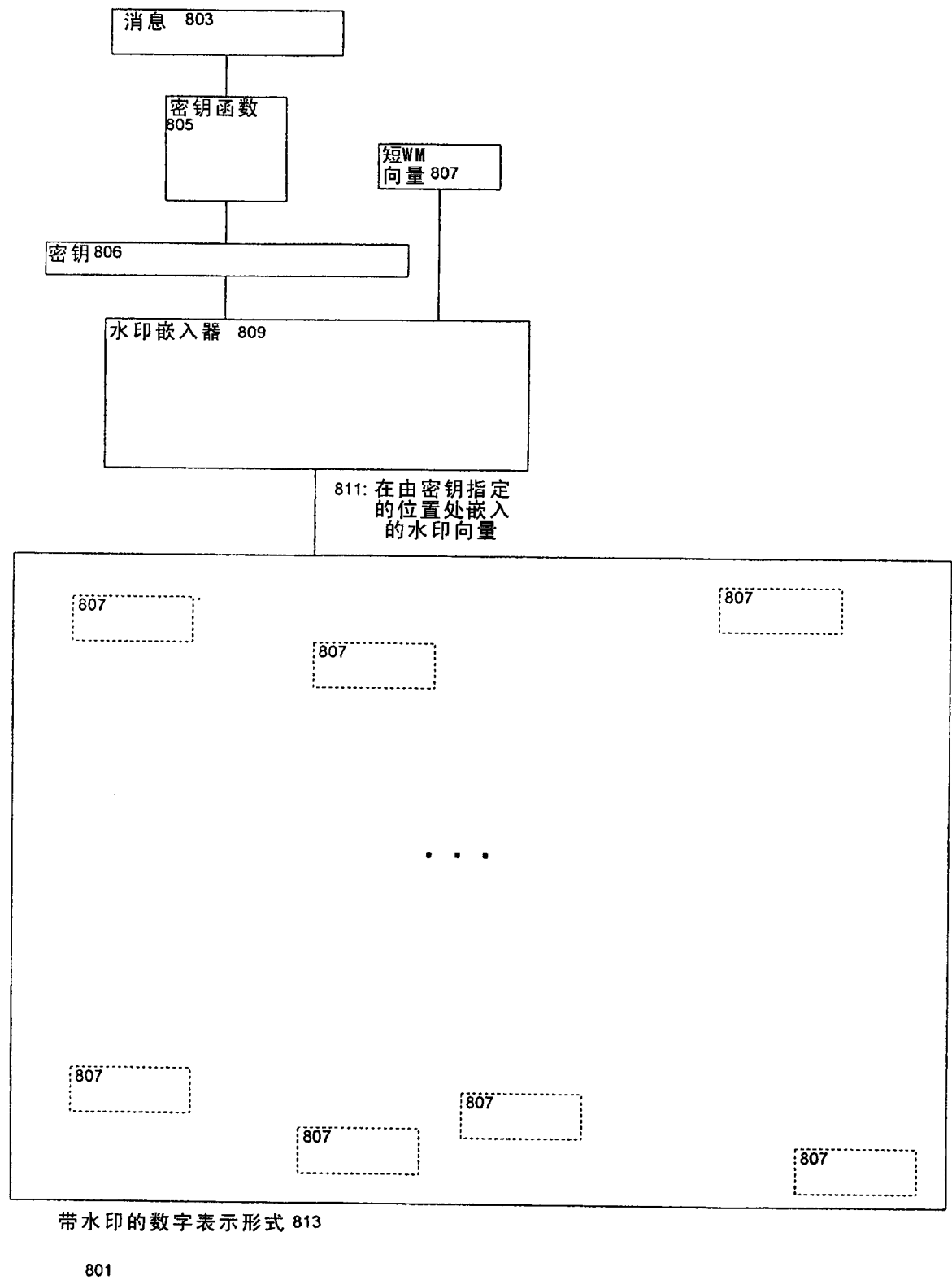
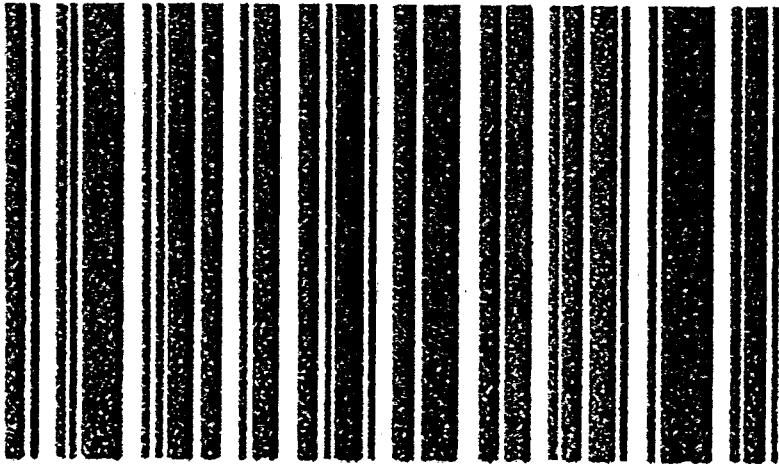
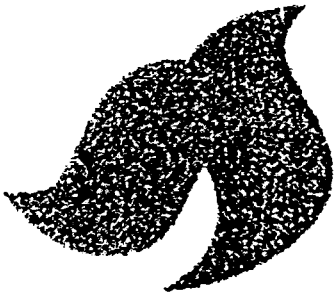


图8

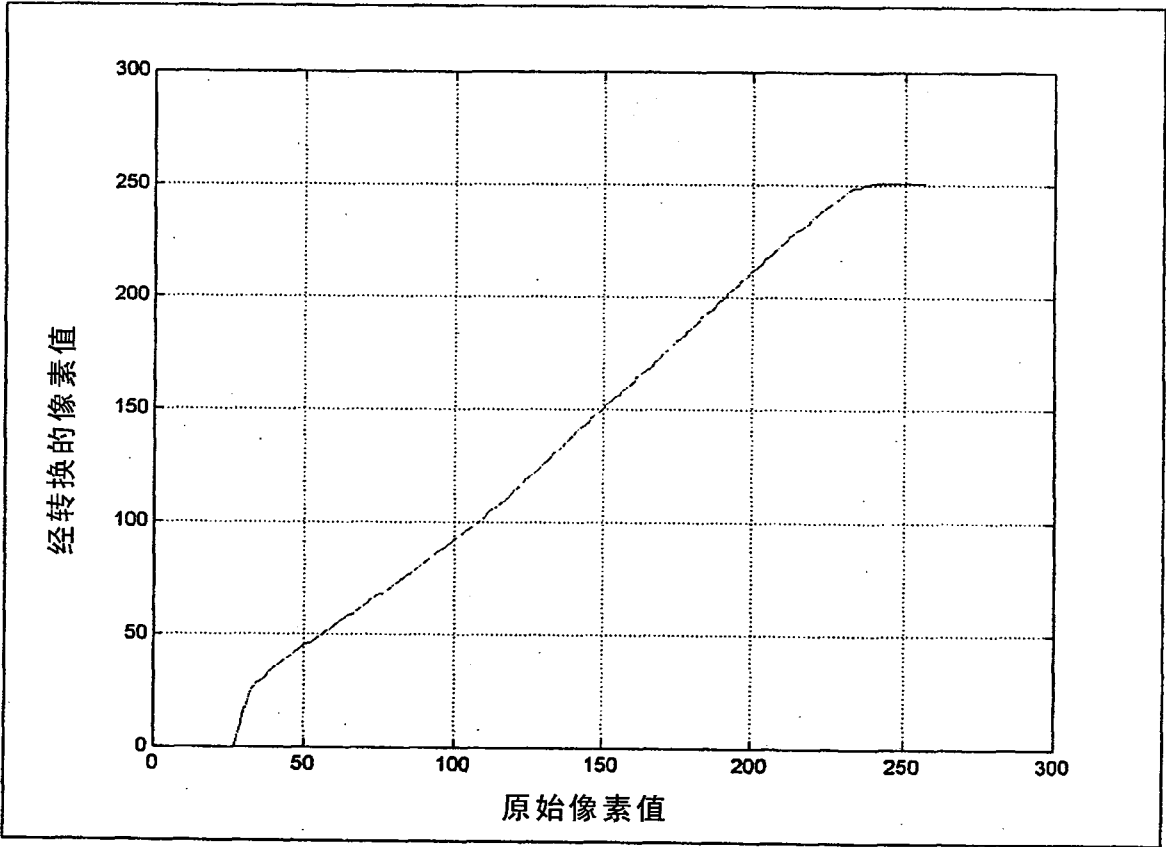


901



903

图9



1001

图10

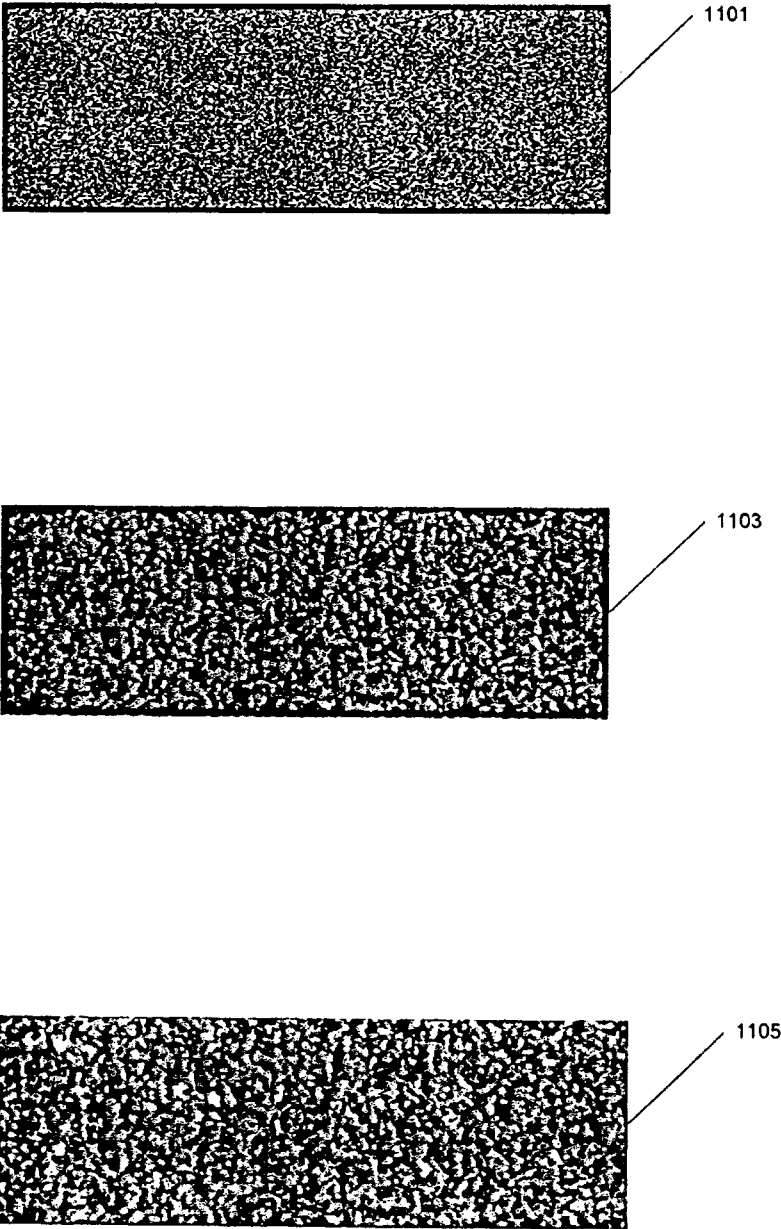


图11

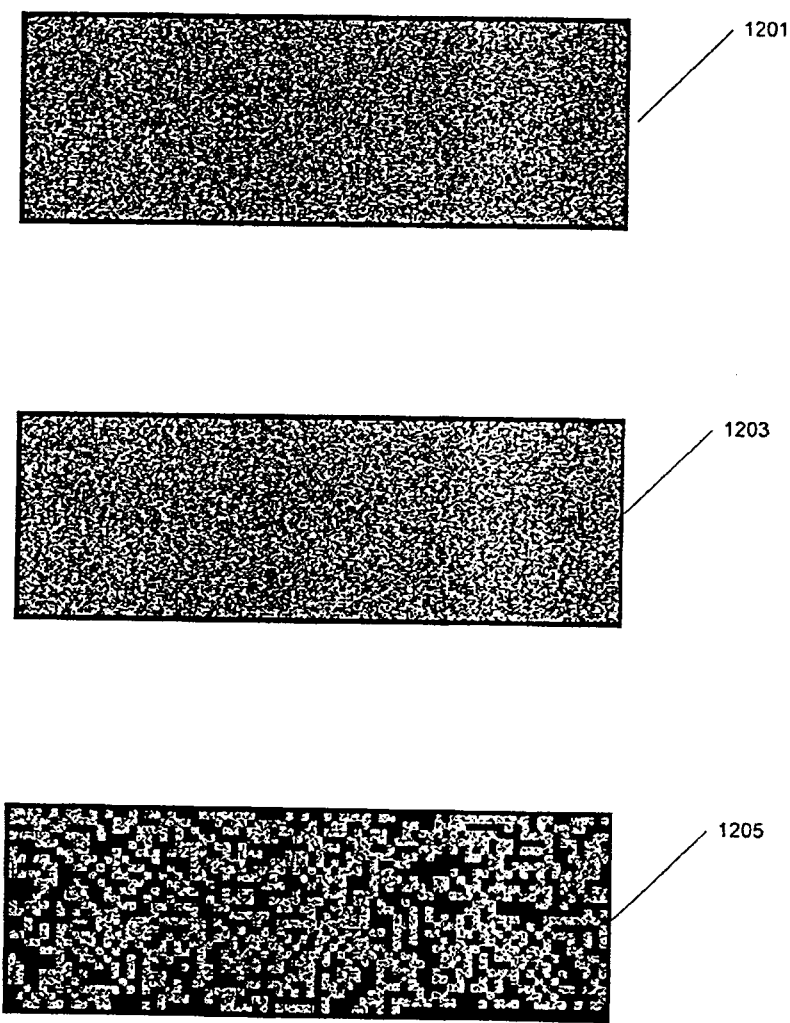
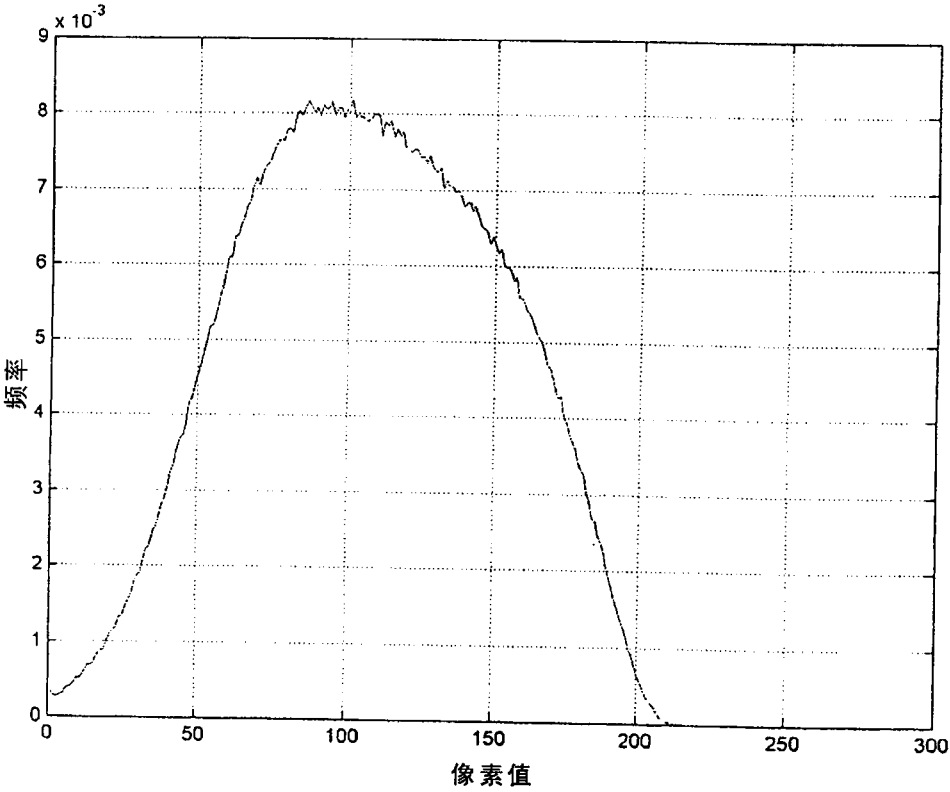


图12



1301

图13

1403

1405(i)

MediSec Technologies LLC
10 Weybosset Street, Suite 501
Providence, RI 02903, USA

Date 04/15/03

1004

Pay to the order of **Bob Smith** \$ **2,234.44**

Two Thousand two hundred thirty four and 44/100 Dollars

Sample Bank

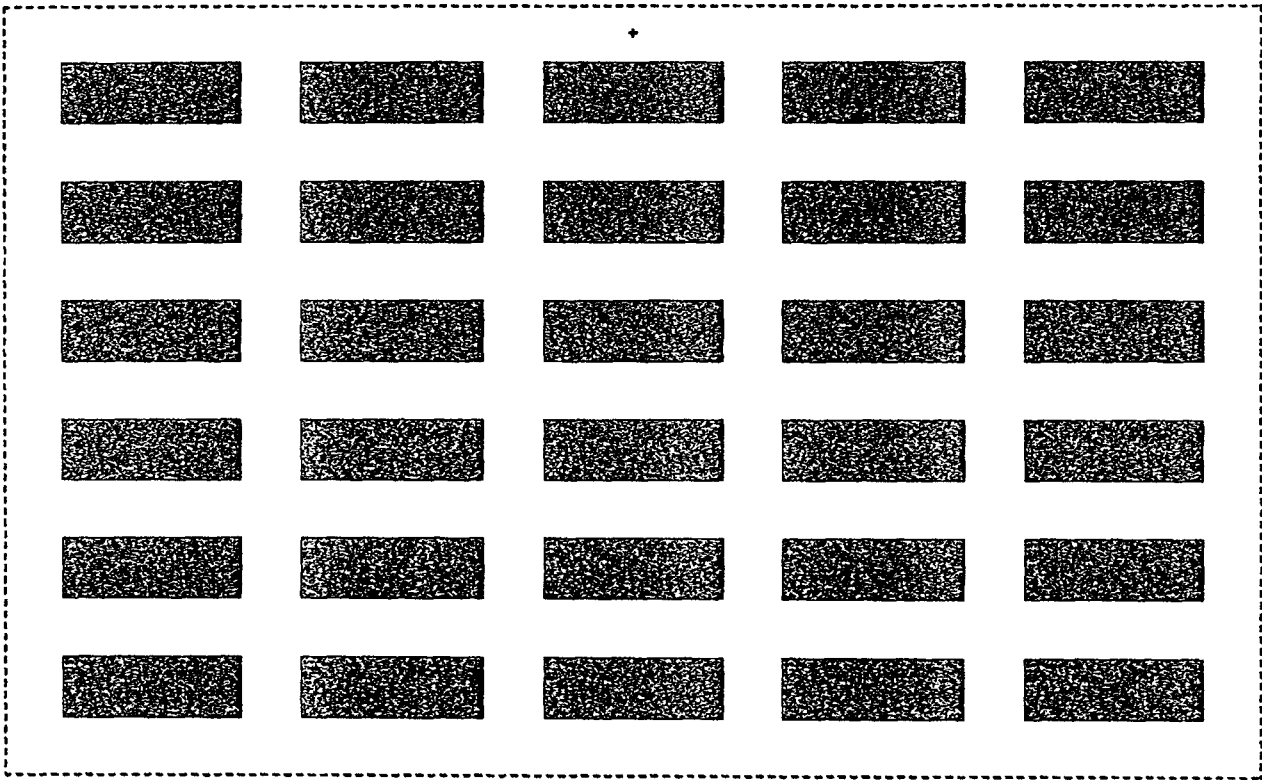
For Sample Sample

1004 1:123451:10041:1234567

demo v2.1 sample #: 00302

1401

图14



1501

图15

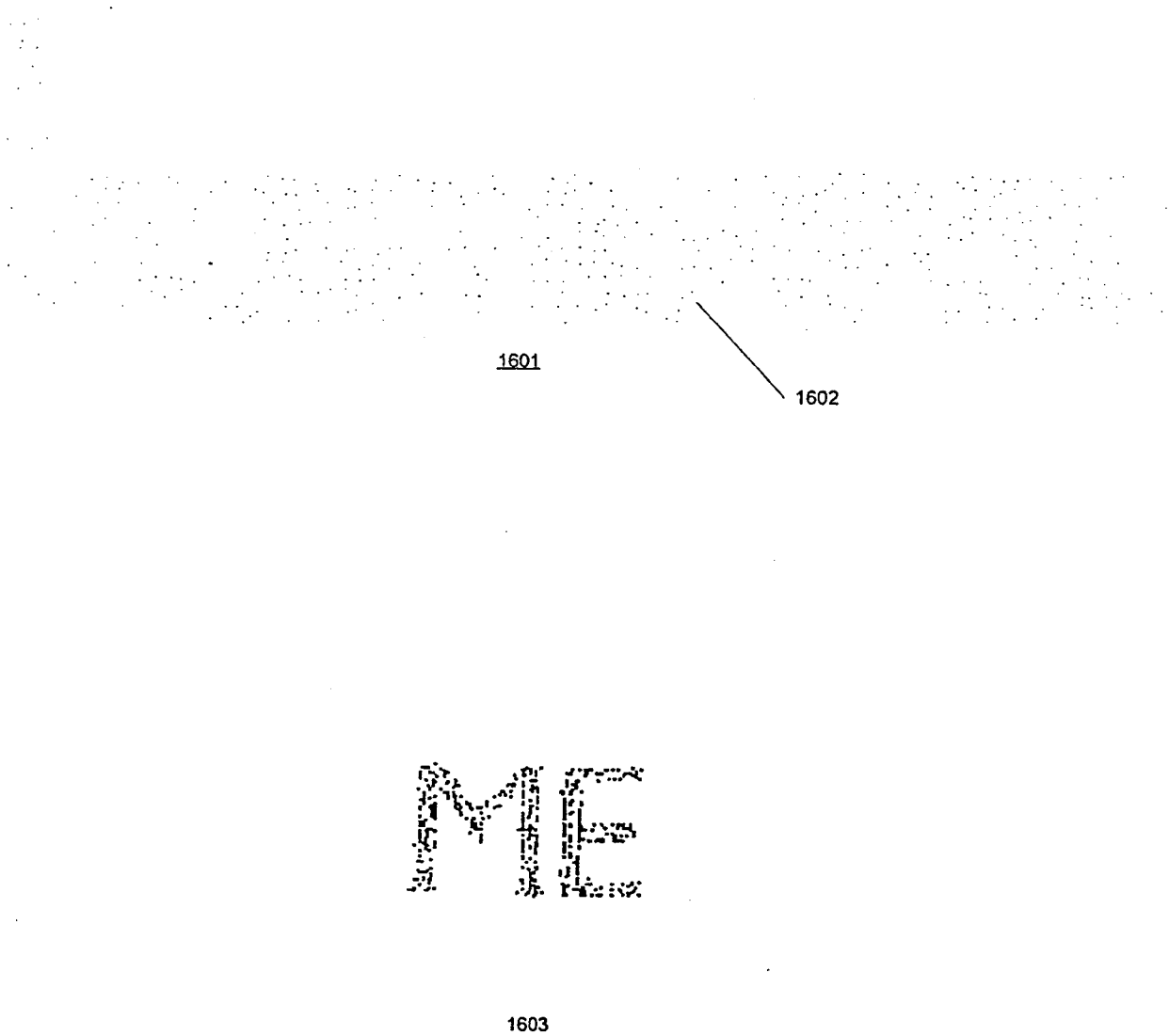


图16

1701

243	228	210	236
59	195	114	189
155	117	158	45
124	5	203	104

1702

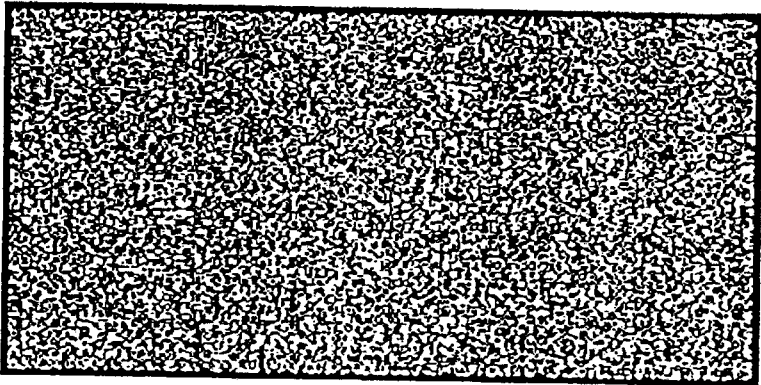
12	27	45	19
196	60	141	66
100	138	97	210
131	250	52	151

1703

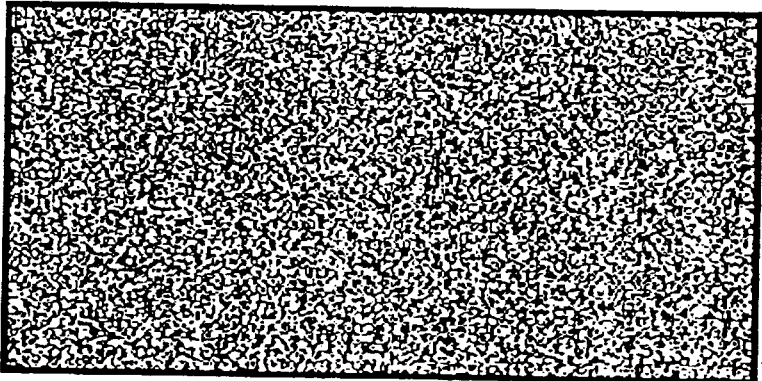
44	36	24	10
198	20	167	83
97	159	135	198
106	299	10	172

图17

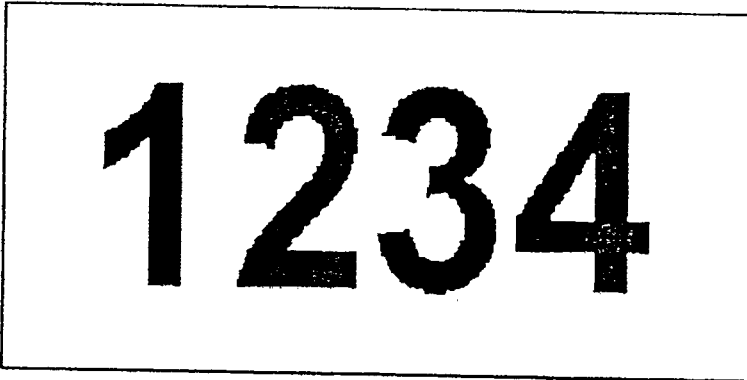
vap0 1801



vapk 1803



VM 1805



1807

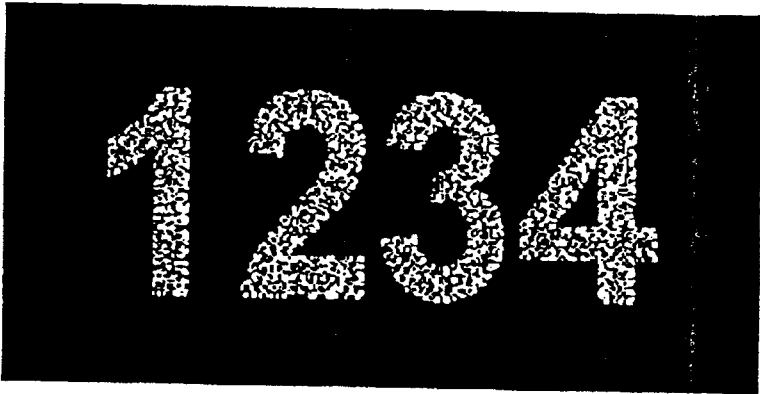


图18