

特辑 2: 新码的提案

纸文件自动读取系统的构筑

阿波罗日本
顾 泽 苍

BA0906-18 0915-1060/08/¥500论文/UCLS

前言

2005 年，个人信息保护法公布并全面实施后。其中，个人信息的 40%以上，通过纸媒体文件泄露出去。面临这一状况，在针对复合机安全机能的纸文件中，进行信息埋入的研究变得备受瞩目。

近些年来，为了降低纸文件的保存·运用·管理的成本，由国家政策大力引导，正迅速从纸文件向电子化转移。但是因为纸媒体具有直感性、安全性、实证性等优点，所以无论电子化如何发展，纸媒体将长久地存在下去。为了对纸媒体进行高效的利用、有效地向电子化转移，能大量埋入信息的技术变得尤为重要了。

此前，为了对从纸文件向电子化转移进行处理，通常是先把纸文件进行扫描，然后作为图像信息登录。这种方法的问题是不能对图像数据进行编辑以及不能登录数据库。

另外，读取技术的常见作法，是使用文字识别 OCR 技术来自动读取纸文件。OCR 技术可以识别文字模式并对文字码进行转换。但是，因为不能获得文档的组版信息和传票的数据库格式等信息，必然需要人的介入。

为了解决上述问题，把纸文件数据作为附加信息，然后以底纹的形式埋入到纸里，要是可以和文件数据一起印刷时，把埋有大量数字印刷物作为记录媒体，可以随时登陆到计算机。还有，纸文件和信息一体化后，优

点是无需多次输入以及登陆前不需要人的介入，可以实现完全的自动读取。而且，纸文件作为记录媒体，要是可以埋入大量的信息，不仅能提高纸媒体的利用效率而且还能减少纸的用量，起到环保的作用。

历来的纸媒体信息埋入码

调查纸媒体秘密文件的流通过程时，因为可以作为数十左右容量埋入，所以历来使用水印技术。这种技术是基于信号处理理论通过设定多个不同搬送方向的特征标记来记述信息。通过信息记录特征标记的打点模式的矩阵配置作成底纹并作为纸文件的背景来印刷。为了隐藏信息记录的特征标记，在信息记录的特征标记的周围需要设定大量没有关系的点。根据这些点的设定，就是出现杂音的原因，对于信息记录特征标记而言会出现误识别的问题。另外，因为是无差别地设定信息记录的特征标记，所以和当纸文件重叠时，有时就不能识别信息记录特征标记。因此，作为同样的信息，需要多次反复地埋入，埋入实用的 100 个字节的信息量。

表 1

	禁止拷贝	追踪功能	防止篡改	自动读取
埋入信息量 (以上)	2Bit	30Byte	500Byte	3KByte

表 1 是针对复合机自动读取机能和安全机能的信息埋入容量的比较。根据表 1 所示，在历来的技术中，可以实现埋入量很少的秘密文件的拷贝禁止和泄漏途径的追踪。但是，不能很好地实现大容量的篡改防止和自动读取等机能。



新纸媒体信息埋入码

在此，和历来水印技术完全不一样，加入了印刷的网屏理论，提出了兼有信息记录码和信息埋入码的新码，即网屏编码。网屏编码对印刷物来说不会产生违和感，就是利用了网屏的特征。首先，如图 1 所示，对于印刷网屏的各个网点，改变各个网点的形式，就可以记述大量信息。因为改变形状的网点的阶调值没有发生变化，所以不会损害画质。

对于纸媒体的文件，埋入大量信息时，底纹方式是最有效的。图 2 的 (a) 是通过网点的大小表示阶调的振幅变调方式，即以 AM 网屏为特征标记“1”的例子。

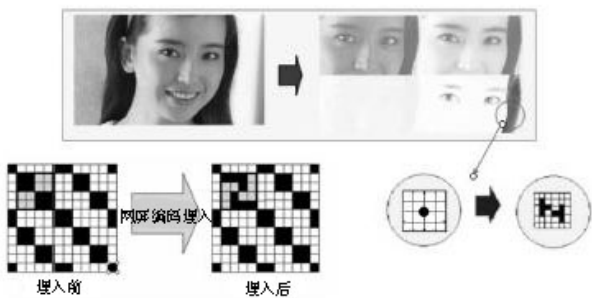


图 1

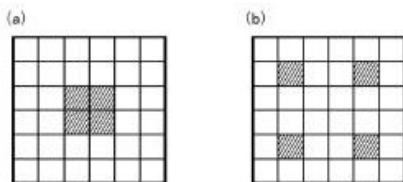


图 2

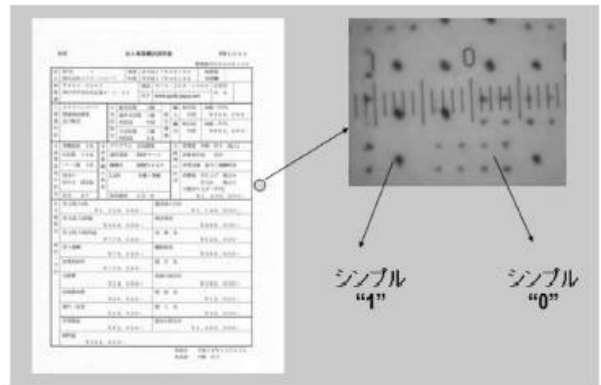


图 3

图 4 的 (b) 是通过网点的点数表示阶调的周波数变调方式即以 FM 网屏为特征标记“0”。即可以通过不同的变调方式记述信息。要是以不同的变调方式的网点为矩阵设定的话，可以组成底纹。另外，即使不同的变调方式，各网点的阶调值不变，所以可能成为一定阶调的底纹。

图 3 是根据不同的变调方式，在纸媒体的文件上埋入信息的例子。如图 3 所示，根据不同变调方式所构成的底纹作为纸传票的背景而印刷的。扩大之后，通过振幅变调方式（AM 网屏）之一的大点构成的网点作为特征标记“1”，由周波数变调方式（FM 网屏）的四个分散了的小点构成的网点作为特征标记“0”。另外，由四个分散的小点构成的网点的大小相加，和一个大网点的点的大小一样，所以要是根据印刷网屏的特点，以一定间隔设定这些不同变调方式的网点，就可以构成均一的底纹。因此没有纸传票的违和感。另外，可以把构成底纹的所有网点作为信息记录的特征标记来埋入信息，所以可以埋入数千个字节以上的大量信息。

在图 4 中，显示网屏编码的处理流程。

从数据库印刷文件和传票等的电子数据时，作为附加信息把纸文件

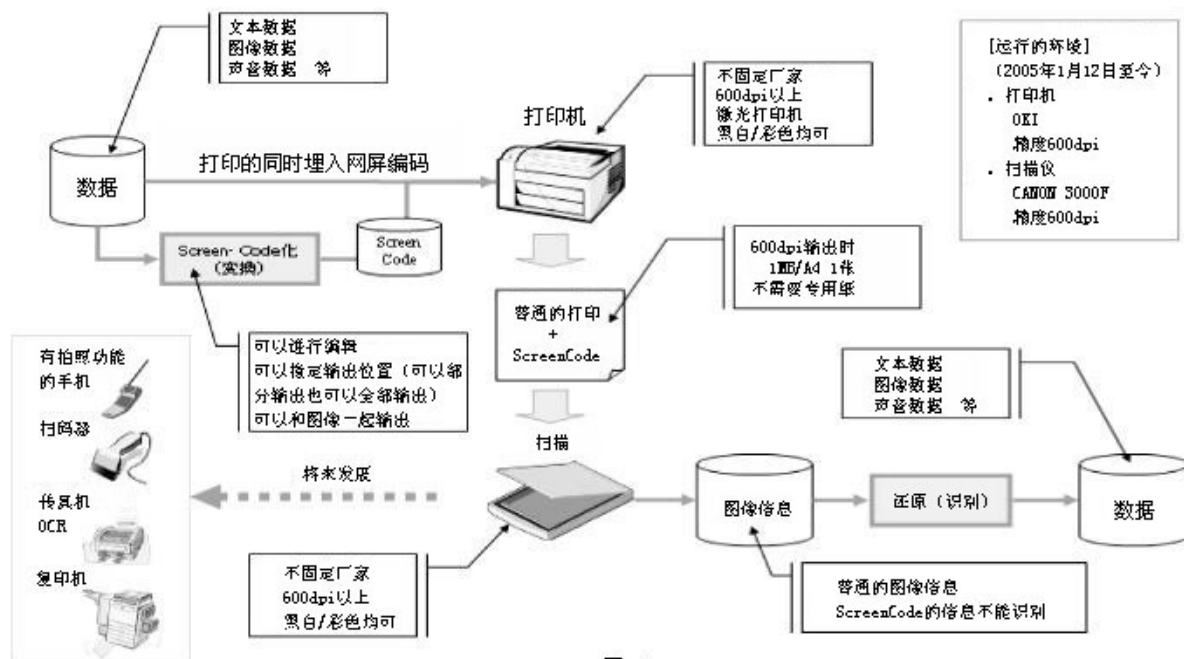


表 2

注：表中（1）和（2）是因为文字识别技术不存在情报记录码，所以没有比较。
表中（3）是因为技术埋入的信息量少，所以自动读取功能不适用。
（4）也是如此。（5）1/3之内的破坏也存在识别的可能。

所要埋入的诊断书数据的选择方法如图 7 所示。

使用网屏编码打印机的设定画面，检查文件的选择，点击参照选择就可以选择埋入诊断书的文件。最近，微软公司宣布公开 Office 软件的格式，所以，今后在使用 Word 和 Excel 等应用软件时，不需要这个设定就可以自动埋入印刷数据。



图 5



图 6



图 7

实际印刷打印机的选择方法如图 8 所示，可以选择实际输出的打印机。

当使用网路打印机时，可以使用上述方法设定。

在读取上述可自动读取的诊断书时，要是使用图 9 所示的一样的软件，就可以自动地把埋入诊断书数据再次登录到计算机中。IC 芯片的方式时，根据网屏编码就可以把信息埋入处理和认识处理 IC 芯片化。如图 10 所示。

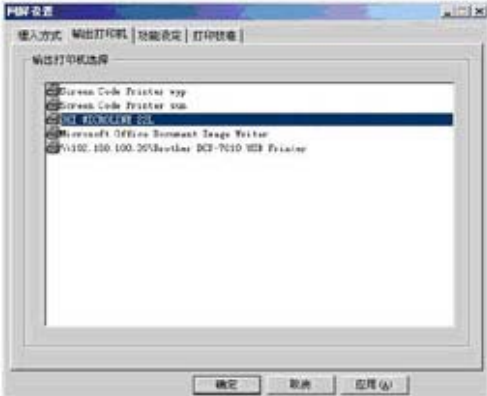


图 8



图 9

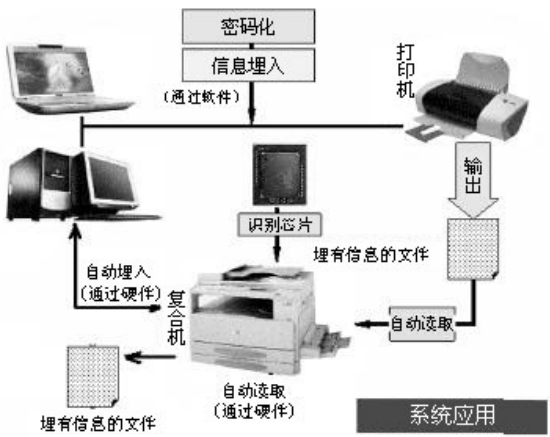


图 10

综上所述，打印机印刷的自动读取可能的诊断书一样的纸文件使用复合机中的扫描仪。对于大量文件而言，可以把埋入的数据连续且高速地登录到计算机中。另外，复合机所印刷的纸文件因为埋入了文件数据，所以根据需要可随时登录到计算机中。



OA 系统的革命

超出常识的自动读取系统的创新，对于 OA 系统而言，可能是革命性的进步。

首先，使用带有自动读取机能的复合机，通过扫描仪读取埋有信息的纸文件时，不仅仅纸文件的表面印刷内容，埋入的信息也可以印刷。如此说来，当从纸文件向电子化处理时，作为可再编辑一样的数据形式，不需要人的介入，可以以完全自动化的方式登录计算机。

一张 A4 大小的纸，可以埋入一本小说的内容，所以要是复合机显示屏的画面比现够大，因为数百张以上的数据埋入到了纸文件中了，所以通过使用大画面的显示器选择必要的文件，就可以进行印刷。另外，对于埋入纸文件的信息，通过密码键就可以控制读的人、场所、时间等。

进而，当制作证明书时，可以把证明内容全都埋入到文件中，所以即使是篡改证明书的表面内容，也不能篡改埋入的内容，所以就可以制作成不可完全篡改的证明书。

所印刷的纸文件，何时、所用的计算机、所用的打印机和文件内容等都可以作为追踪信息而埋入，纸文件流通的各个阶段都可以追踪到，这样就可以提高纸文件的泄露管理。另外，通过上述的追踪信息，也可以认证纸文件的记时打印等。

自动读取系统統合了纸媒体的直感性、廉价性和证明性等特征以及电子媒体的详细性、检索性和便利性等特征而生产出高效率的产品。另外，把大量信息埋入纸文件，可以大幅度地削减纸的使用量，可以说是一种

环保的技术。

对于文档和商务表格的应用程序系统而言，历来只能作成书面的表面数据。从今后，在纸文件里埋入信息的处理流程和数据格式的统一等将是今后的重大课题。



今后的发展

基于印刷网屏的特性所构成的网屏编码，根据打印机的印字精度的提高，应用范围也将飞跃地扩大。对于以前认为不可能的和此前不敢想的问题，现在可以开发出许多创新和新的综合解决方案。由印刷基础技术所开发的网屏编码，对于纸文件而言可以说是划时代的技术革新。

笔者介绍

顾 泽苍

阿波罗日本公司

董事长 工学博士

日本电子信息通信学会 (IEICE) 和图像电子学会 (IEEEJ) 的正式成员

株式会社 阿波罗日本

〈负责人〉 董事长 顾 泽苍

〈本公司住址〉

中国天津市南开区华苑产业园区物华道2号火炬创业园
A座4楼494

URL : <http://www.apollo-japan.ne.jp>

E-mail : info@apollo-japan.ne.jp

〈事业内容以及公司近况〉

本公司独立开发的网屏编码，在日本已经申请了两个专利，对于导入到安全领域获得了日本某著名大公司的日本顶尖水平的技术评价。

利用业界首创的可以在纸上埋入大量数据这个优越性，瞄准从纸文件向电子化转变的庞大的商业市场，开发带有自动读取机能的复合机，以日本大的复合机生产厂家为对象，拓展认证销售业务。