

## 単一ドットパターンの抽出法改善と前景耐性 向上の検討

橘義紋<sup>†</sup> 小野要<sup>††</sup> 金田北洋<sup>†††</sup> 岩村恵市<sup>††††</sup>

近年印刷物からの情報漏洩が増えているため、その情報保護が重要になってきている。そこで我々は、視覚品質や埋め込み情報量に関して利点がある単一ドットパターンを用いて情報を埋め込む手法を研究してきた。しかしながら、従来研究では紙面の微少なゆがみの問題等から A4 1 ページ全領域からの抽出は困難であった。そこで、本研究では抽出法を改良し、A4 1 ページ全領域からの抽出を可能にした。さらに、OCR 技術を用いた前景耐性手法をデータと領域を拡げてその効果を定量的に把握した。

### Study on improvement of extraction of simple dot pattern and robustness to the foreground

YOSHIAKI TACHIBANA<sup>†</sup> KANAME ONO<sup>††</sup>  
KITAHIRO KANEDA<sup>†††</sup> KEIICHI IWAMURA<sup>††††</sup>

Because more recent information leak from the print has become important to protect that information. Therefore, we have studied the method to embed information using simple dot pattern has the advantage for the amount of embedded information and visual quality. However, because of previous problems and slight distortion from all areas of the paper on page A41 of the extraction was difficult. Therefore, this study is to improve the extraction, A41 made it possible to extract from all areas of the page. In addition, OCR was quantitatively assess the effect of data and methods to expand the area of technical robustness for the foreground.

## 1. 背景\*

近年プリンタの印字解像度の向上 (1200dpi, 2400dpi 等) により、一般的に普及している安価なプリンタでも高品質の文字を形成することが可能となり、印刷物に対する不正コピーや改竄がより手軽かつ巧妙に実現できる環境にある。実際、最もインシデント件数が多い漏洩媒体・経路は印刷物で、全体の 73% となっている [1]。紙媒体は、業種や業務内容に関わらず、使用機会の多い媒体であるため、その抑止技術としての印刷物への情報埋込み・抽出技術は重要性が増している。

現状の印刷物への情報埋込み・抽出技術は大別すると紙に印刷する画像や文字などの電子コンテンツに直接電子透かしを埋め込む手法 (不可視) と、紙にバーコードや特殊なパターン等を印刷する手法 (可視) の 2 つに分類される。これらはそれぞれ長所と短所があり、情報量・処理速度・精度・コストのすべての要求を満たす方式は今のところ存在していない。

我々は従来手法の課題であった文書品質の劣化、埋込情報量のコンテンツ依存性などを解決するために視覚品質や埋め込み情報量に関して利点がある単一ドットパターン [2] を考案し、このパターンを用いた印刷文書に対する新しい情報付加手法を提案している。

しかしながらこれまでの単一ドットパターンの抽出は、紙面の微少なゆがみなどの問題から、名刺サイズの領域で実証実験を行っていた。本研究では実用性を高めるために、紙面の微少なゆがみに影響されない抽出法を検討し、結果として A4 1 ページ全領域からの単一ドット抽出を実現可能にした。さらに、OCR (Optical Character Recognition) 技術を用いた前景耐性手法をデータと領域を拡げてその効果を定量的に把握し、今後の課題を明確にした。

## 2. 従来手法

### 2.1 沖電気方式

2004 年に沖電気工業社から「印刷文書への透かし埋込および検出方法」という題目で、印刷物の背景に情報を埋め込む方式の論文が発表された [3]。この方式は、18×18 画素を単位とする網掛けパターンを背景にタイル状に並べていく。図 1 にこの網掛けパターンを示す。ドットが等間隔に配置されているこの網掛けパタ

\*<sup>†</sup> 東京理科大学大学院  
Tokyo University of Science Graduate School  
<sup>††</sup> (株)野村総合研究所  
Nomura Research Institute, Ltd.  
<sup>†††</sup> (株)キヤノン  
Canon Inc.  
<sup>††††</sup> 東京理科大学  
Tokyo University of Science

ーンを基本網掛けパターンと呼び、このパターンを通常は埋め込んでいく。そして何か情報を埋め込みたい場所には、基本網掛けパターンの一部分をずらした“0”か“1”のパターンを埋め込むというものである。図2に“0”と“1”の網掛けパターンを示す。この論文ではプリンタの解像度を600dpiとしている。この解像度ならドットのパターンが人の目には見えにくく、背景として違和感が無いだろうと考えられるためである。

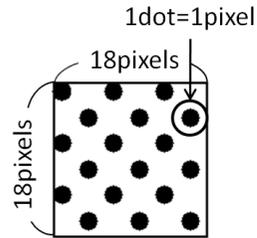


図1 沖電気工業方式の基本網掛けパターン

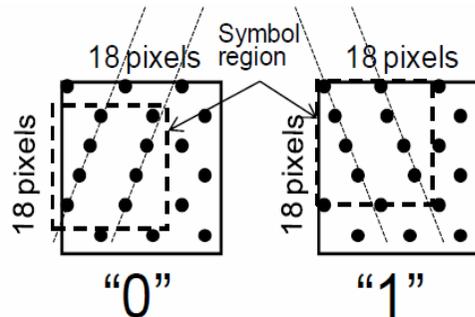


図2 沖電気工業方式の“0”と“1”の網掛けパターン

パターンを抽出するには紙面をスキャナーで読み取り画像ファイルを生成するが、印刷→スキャンという工程でドットが明瞭でなくなっていることがある。ちなみにスキャナーの解像度は400dpiとしているので、なおさらドットが粗くなっている可能性は高いであろう。そのため直接パターンを検出するのではなく、図2の斜め線によって決定される周波数と方向を持つ波を検出する。波の方向の区別には透かし検出フィルタとして二次元ガボールフィルタを用いる。ガボールフィルタは窓関数にガウス関数を用いており、ノイズの多い検査画像に於いても波の方向を安定して検出することが可能である。

### 2.1.1 沖電気の方式の問題点

実際に図2に示す網掛けパターンを用いて情報を埋め込んだ紙を図3に示す。また、図3の紙に文章を印刷したものを図4に示す。

この網掛けパターンでは18×18画素という約0.5mm四方の微小な空間にドットを18個も用いるため、図3、4から分かるように解像度が高くて、背景濃度が上昇し、全体がグレートーンとなり、文書品質の劣化を招いてしまう。さらに、抽出する際にガボールフィルタを用いているが、ガボールフィルタはその計算量の多さで複雑なプログラムを要するため、処理速度に対する悪影響も少なからずあるとも考えられる。

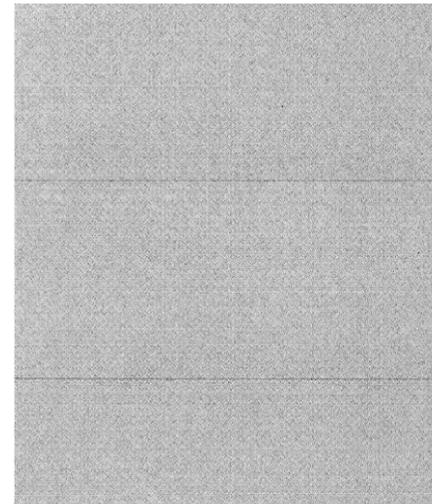


図3 沖電気の方式を埋め込んだ紙

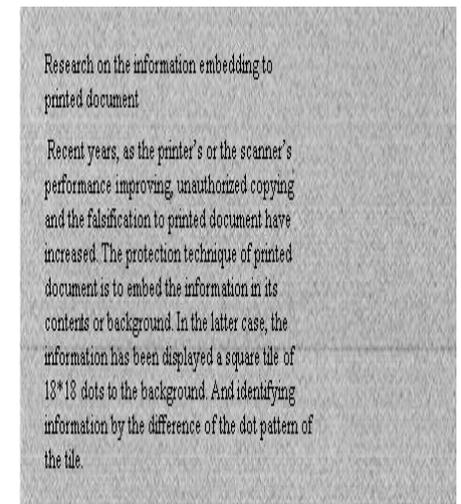


図4 沖電気の方式に文章を印刷した紙

### 2.2 単一ドットパターン

そこで、上記の背景に情報を埋め込む手法の課題点である文書品質の低下を克服し視覚的な文書品質の向上を目指し、ガボールフィルタを用いずにより簡単に埋め込まれた情報を検出することを目的として提案された方式が、2008年に発表された「単一ドットパターンを用いた印刷文書用電子透かしに関する一提案」である[2]。

この方式は沖電気の方式と同様に、文書画像の背景に情報を表現する網掛けパターンを重ね合わせた方式である。網掛けパターンの構成としても沖電気の方式と同様に、18×18画素を単位とする網掛けパターンを背景にタイル状に並べることによって生成する。図5に示すように“0”と“1”の情報表現の仕方は、18×18画素の中にドットを1つ打ったものを“1”の網掛けパターン、何も打たないものを“0”の網掛けパ

ターンとして情報を埋め込む。

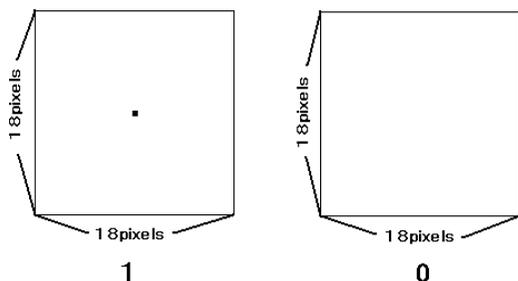


図5 単一ドット方式の“0”と“1”の網掛けパターン

ちなみにプリンタとスキャナーの解像度は600dpiとしているが、600dpiでは1ドットの大きさが0.042mmで、点と点の間隔が18ドット(0.75mm)となり、視覚的に違和感を覚えにくいと考えられるためである。また、スキャンする際に紙が斜めになっていたり折り曲がったりしていないことを前提条件としている。実際に図5に示す網掛けパターンを用いて情報を埋め込んだ紙を図6に示す。また、図6の紙に文章を印刷したものを図7に示す。

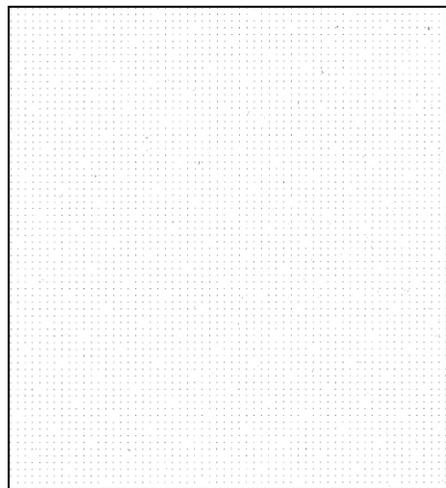


図6 単一ドット方式を埋め込んだ紙

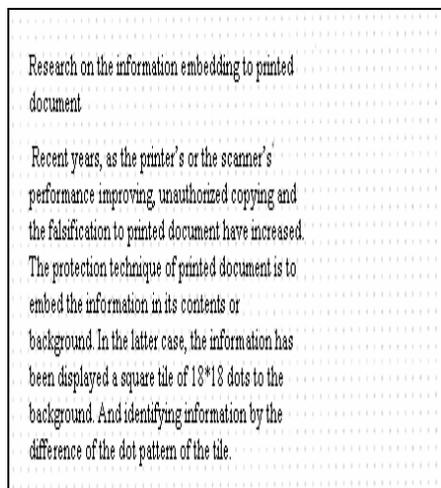


図7 単一ドット方式に文章を印刷した紙

18×18画素という約0.5mm四方の微小な空間にドットを18個も用いる沖電気の方式と比べて、単一ドット方式はドットを1個用いるだけなので図3、4と図6、7を見て分かるように、視覚的な文書品質において沖電気の方式よりも優れていることが

分かる。またこの手法では、情報を1ドットの有無で表しているの、抽出は基本的な加算・比較演算の組み合わせで実現でき、計算量においてもガボールフィルタを用いる沖電気の方式よりも簡易であることが分かる。

### 2.2.1 これまでの成果

これまで単一ドットに対して行われてきた研究では、18×18画素の網掛けパターンを9×9画素および7×7画素に縮小したパターンで埋め込み抽出率がどのようになるかの検証を行って、いずれのパターンサイズにおいても90%以上の抽出率であった。また名刺サイズ(2160×1296画素、600dpi)に情報「1010」を埋め込むことを前景ありと前景なしで行った。前景ありではOCR技術を用いて文字部分を避けて情報を埋め込み抽出の検証を行い、前景なしと前景ありは共に98%前後の抽出率を得ている[4]。

## 3. 実験

### (ア)実験の目的

本研究では、従来行われていた前景ありと前景なしのサイズを一般的に使用されることの多いA4サイズに拡大する。さらに、今までは文字部分を文章または段落単位の塊として扱い、それを避けて情報を埋め込んでいたので、ドット部分だけを切り取ることは容易であり、その場合情報漏洩を抑止することができなくなってしまうという問題点があった。そこで、ドット部分だけを切り取りにくいように文字と文字の間にも情報の埋め込みを行うことで抽出率がどのようになるかの検証も併せて行う。

### (イ)使用機材

今回の実験で使用した機材は以下である。

- A) プリンタ・Canon製 MP970 (印刷解像度 600dpi)
- B) スキャナー・Canon製 MP970 (読取解像度 600dpi)
- C) 使用コンピュータ、及びソフト

◎PC: HP xw4550 Workstation Dual-Core AMD Opteron(tm) Processor 1212 2.00GHz, 896MB RAM ◎OCR: 読取革命 Ver.12 ◎開発環境: Microsoft Visual C++ 6.0

### (ウ) 実験方法

まず従来行われていた抽出範囲の名刺サイズからA4サイズに対応する際に生じる紙のゆがみによる問題を検討した。

その結果、抽出範囲が拡大することで紙方向と、搬送方向のズレ、及びドットピッチのズレに影響されて情報を埋め込む際にドットが等間隔で打たれずズレてしまうという問題が判明した。実際は徐々にドットが等間隔でなくなり打ち込むドットがズレていくが、図8にドットのズレをわかりやすく表したものを示す。

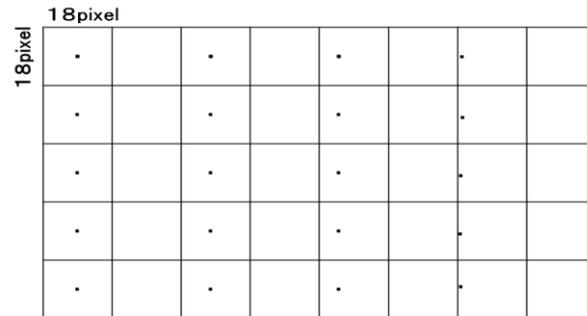


図8 ドットのズレ

図8を見てわかるように18×18サイズのタイルに正しく打ち込まれておらず、ドットがズレてしまう。

ドットズレの原因を以下のように分析した。まず仮定として、①ドットを打ち込むプログラムに誤りが生じて起こる、②プリンタで紙に印刷・スキャナーで紙から抽出する際にドットにずれが生じる、の二つの原因を考えた。1つ目のドットを打ち込むプログラムに誤りがあるか判断するため、情報を印刷物に出力するのではなくBMP形式の画像に直接出力するようにした。このようにすることでプログラムにより指示しているそのものが表示することができ、その画像のドットを見てプログラムによる誤りがあるか判断することができる。BMP画像に出力した結果ドットはズレを起こさず、等間隔で打ち込まれていた。また2つ目のプリンタで情報を埋め込む際に誤りが起こるのか判断するために、印刷物に埋め込みを出力するプリンタを他のプリンタに代えて行った。他のプリンタを用いて出力したのも同じようにドットがズレてしまうことがわかった。これら2つの結果からドットがズレてしまう原因は、プログラムによる誤りでなくプリンタまたはスキャナーによる紙への情報を埋め込み・抽出の際に生じるものだと判断した。A4サイズに打ち込んだ時に生じるドットの全体的なズレをわかりやすく表したものを図9に示す。

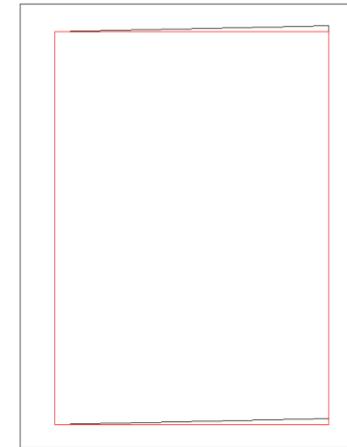


図9 全体的なズレ（黒線→実際、赤線→理想）

図9のA4の紙サイズに対してプログラムによる情報埋め込みは、等間隔で行われ赤線のように長方形に埋め込まれるようにしているが、プリンタまたはスキャナーを用いて印刷物に実際に埋め込み・抽出の際ドットにズレが生じて黒線のように平行四辺形に埋め込まれてしまう。使用しているプリンタやスキャナーは一般的に使用されるものなので、単一ドットの実用性を考えるとドットのズレが生じることに對して情報を抽出する際に對する必要がある。図10に示すドットが打ち込まれている左端aと右端bの位置からab間の傾きを知ること、次にドットが打ち込まれている場所の位置がわかってズレに對するさせることができる。そのためまず、紙のどの位置からどの位置までドットが打ち込まれているか調べるために、紙の左端から順番に横に1画素・縦に1列分の合計画素値を調べた。ドットが打ち込まれていない列つまり空白の列は、画素値が最大の白の値を足していくので合計画素値は最大となるが、ドットが打ち込まれている列になるとドットがある分、黒に近くなり画素値が下がることで合計画素値も下がるので、閾値を設けて合計画素値が閾値を下回った列をドットが打ち込まれた左端の部分と判断した。同様に、ドットが打ち込まれた上端・右端・下端を調べた。調べた上端と右端から図10のbの位置は判明するが、aの位置は図10に示すように上端ではないため上端と左端から判明することができない。そこで、aの位置は打ち込まれたドットの行数×18pixelの計算を行い、下端にその分を足してaの縦方向の位置を出し左端と合わせてaの位置を判明した。この判明したaとbの位置から傾きを出し、抽出する際にはこの傾きに沿って抽出する位置を変えてズレに對するさせた。2行目以降は同じ傾きを前提として処理を行った。



図 10 図 9 の拡大図 (黒線→実際、赤線→理想)

次にドット部分だけを削除することに耐性を持たせるために、文字と文字の隙間にドットを埋め込み抽出を行うためにプログラムを作成した。従来は OCR を用いて OCR アプリから出力される文字位置・認識結果等の XML ファイルのデータによって文字部分を文章または段落単位の塊として扱い、それを避けて情報を埋め込んでいたので、今回は OCR を用いて OCR アプリから出力される文字位置・認識結果等の XML ファイルのデータによって文字部分を判断し避けていた部分の文字と文字の隙間に情報を埋め込むようにした。抽出の際にも OCR アプリから出力される文字位置・認識結果等の XML ファイルのデータを用いて文字部分を判断して抽出を行った。文字間にドットを埋め込むプログラム処理を図 11 に示す。

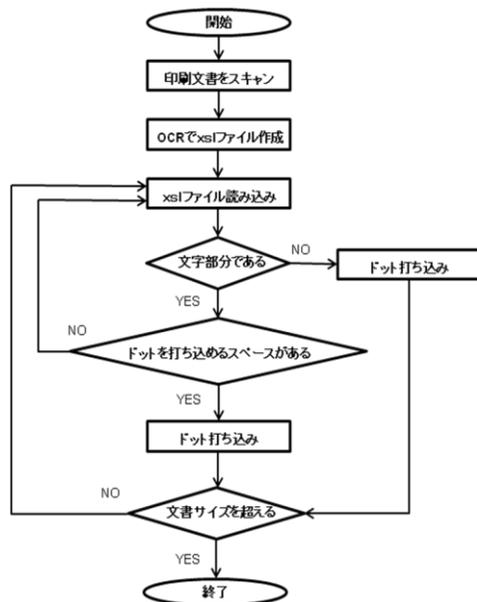


図 11 文字間ドットを埋め込む処理

抽出の時は図 11 の打ち込む処理の流れと同じようにして行き、[ドット打ち込み]

のところを抽出にして抽出処理を行うようにした。

#### 4. 結果

情報埋め込む範囲を名刺サイズから A4 サイズに拡大して抽出を行った結果を表 1 に示す。また今回使用したサンプルを図 12 [5] に示す。

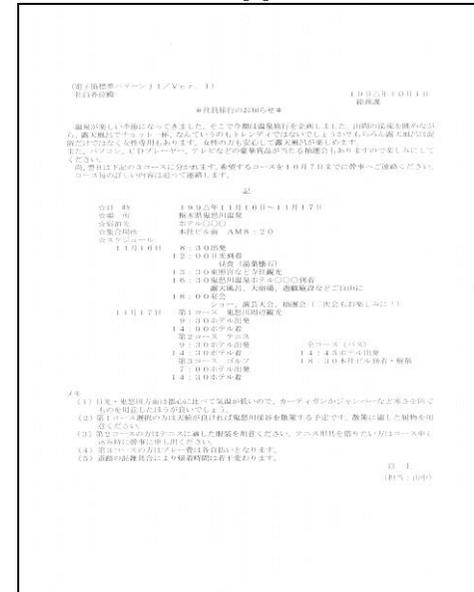


図 12 使用した原稿サンプル

表 1 サイズ拡大

	名刺サイズ		A4 サイズ	
	なし	あり	なし	あり
前景	なし	あり	なし	あり
埋込可能情報量[bit]	8064	7450	101572	88963
抽出失敗情報量[bit]	189	108	3827	17296
抽出率[%]	97.68	98.55	96.23	80.56

表 1 より名刺サイズから A4 サイズに拡大すると前景なしとありの時どちらも抽出率が下がっていることがわかる。また前景ありの時に 80%にまで抽出率が下がっている原因としては、印刷スキャン時の紙のズレによって文字認識に微妙な誤りが生じたため前景部分をうまく避けて抽出が行えなかった箇所があると考えられる。

次に、文字間にドットを埋め込み抽出を行った結果を表 2 に示す。また、文字間にドットを埋め込んだ拡大図を図 13 に示す。

☆日 時	199△年11月16日～11月
☆場 所	栃木県鬼怒川温泉
☆宿泊先	ホテル○○○
☆集合場所	本社ビル前 AM8:20
☆スケジュール	
11月16日	8:30 出発
	12:00 日光到着
	昼食 (湯葉懐石)
	13:30 東照宮など寺社観光
	16:30 鬼怒川温泉ホテル○○○
	露天風呂、大浴場、遊戯
	18:00 宴会
	ショー、演芸大会、抽選

図 13 文字間埋め込みの拡大図

表 2 文字間にドットを埋め込み

	文字間
埋込可能 情報量[bit]	91615
抽出失敗 情報量[bit]	18110
抽出率[%]	80.23

表 2 より文字間にドットを埋め込むことで埋め込み可能情報量が増えていることがわかる。しかし抽出率も少し下がっていることもわかる。その原因として情報を埋め込むスペースがあるにも関わらずドットが打ち込まれていないことや、文字があり打ち込むスペースがないのに判断を誤り、文字にかぶせて情報を埋め込んでしまった箇

所があると考えられる。

## 5. まとめ

本論文では印刷文書に対する情報付加手法の 1 つである単一ドット方式を用いて埋め込む領域を拡大し情報量の向上させることができた。また、前景部分の切り取り耐性の向上のため OCR アプリから出力される文字位置・認識結果等の XML ファイルのデータを用いて情報の埋め込み及び抽出を行った。

その結果埋め込む領域を拡大し情報量を増やすことができ、80%の抽出率を得た。また、前景部分の切り取り耐性のために文字間にドットを埋め込むことが可能となり、視覚的にも違和感の少ない印刷物を作成でき抽出率も 80%を保つ事ができた。埋め込む領域を拡大し文字間にドットを埋め込むことで抽出率が 80%まで下がった原因としては、印刷スキャン時の紙のズレによって文字認識に微妙な誤りが生じたため前景部分をうまく避けて抽出が行えなかった箇所があるため下がってしまった。また、文字以外のグラフや図などに対しては前景部を正しく判断することができず、文字認識に誤りが生じるため正しく情報を抽出できないことがわかった。

今後は抽出率の向上や文字以外のグラフや図などの前景部に対しても対応していくようにし、実用化を目指すための検証実験を行っていきたい。

謝辞 本報告作成のためにご協力頂いた岩村研究室の皆様に、謹んで感謝の意を表す。

## 参考文献

- 1) NPO 日本ネットワークセキュリティ協会、“2009 年情報セキュリティインシデントに関する調査報告書”
- 2) 金田北洋、永井文也、岩村恵市、半谷精一郎、“単一ドットパターンを用いた印刷文書用電子透かしに関する一提案” 第 4 回マルチメディア情報ハイディング研究会、pp11-16、July 2008、copyright(c) 2008 IEICE
- 3) 須崎昌彦、須藤正行、“印刷文書への透かし埋込及び検出方法” 進学論 (A)、vol.J87-A、no.6、pp778-786、June 2004
- 4) 小野要、李柱昊、金田北洋、岩村恵市、“単一ドットを用いた情報付加手法の文字印刷耐性に関する研究” 情報処理学会研究報告 IPSJ SIG Technical Report 2009 Information Processing Society of Japan
- 5) プリンタ専門委員会 (技術分科会) 情報処理標準化運営委員会 “IT-3011 プリンタ用標準テストパターン 2003 年 3 月制定” JEITA 電子情報技術産業協会規格 社団法人 電子情報技術産業協会