

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

小野 要[†] 李柱昊[†] 金田北洋[†] 岩村恵市[†]

概要 印刷物の情報保護手法にはいくつかの技術がある。その中でも単一ドットパターンには、視覚品質と埋め込み情報量に関して利点があるが、この方法には前景が干渉する問題がある。この研究では、単一ドットパターンの小型化と、OCR 技術を用いた前景分離が可能であるかを検討した。その結果、単一ドットパターンの小型化と、視覚品質を保ちながら前景部分の分離が可能であり、かつ 97%以上の情報抽出に成功した。

キーワード 電子透かし, 印刷文書, 単一ドットパターン, OCR

A Study of Information Hiding Performance using Simple Dot Pattern with different Tile sizes

Kaname Ono[†] JooHo Lee[†] Kitahiro Kaneda[†] Keiichi Iwamura[†]

Abstract There are several proposed techniques on the information hiding for printed document. Among other things, Single dot pattern have advantages regarding a visual quality and information capacity. However, this method has serious problem of interference with the foreground set, therefore it is not able to extract information definitely for the document which a letter appears in.

In this research, we examined the extraction and implantation of the information after having extracted a foreground set part by existing OCR technology using simple dot pattern.

As a result, we can bury information avoiding a foreground set part while keeping visual quality, and we acquire over 97% information extraction accuracy.

Keyword Information hiding, Printed document, Single dot pattern, OCR

1. はじめに

近年、一般的に普及しているプリンタやスキャナなどのハード面の進化に加わって、画像編集アプリケーションなどのソフト面の性能の向上により、印刷物に対する不正コピーや改ざんがより簡単に、かつ巧妙に実現できる環境になってきている。実際、

印刷物からの漏えいは情報漏えい事件全体の 46%を占めており、その対策が急務とされている[1]。印刷物への電子透かしとして、不可視の状態で作権情報を埋め込み、透かしの検出を行うことで著作権情報を確認する、という条件を満たすような保護技術の重要性が増している。だが、まだ印刷物を対象とする保護技術は少ないのが現状である。

現状の印刷物への情報埋め込み・抽出技術は大別すると情報を前景（コンテンツ）に埋め込む手法と背景に埋め込む手法に分類される。これらはそれぞれ長所、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ない。

前者は情報を印刷したいデジタルコンテンツの中になるべく目立たないように、かつ印刷しても情報が消失しないような耐性を持たせ、いわゆる電子透かしとして不正コピーなどによる情報漏洩から著作権者の著作権を守る技術として活用されている。しかしながら文書品質の劣化、埋め込み情報量が文字量により左右される、等の欠点がある。

後者は背景パターン、バーコードなどに情報を埋め込む手法に代表される。本方式は埋め込み情報量が多いが、文書デザインの自由度、前景の視認性が劣化するという欠点がある。本研究においては印刷文書の背景に情報を埋め込む方法[1][2]に関して、その課題を克服する新しい情報埋め込み手法を提案し、単一ドット方式の実用化を目指すために既存の OCR 技術（光学文字認識（Optical Character Recognition）の略で、通常スキャナで取り込んだ活字を主とする文書画像をコンピュータが編集できる文字コードの列に変換する技術）を用いて前景部を抽出した後抽出する手法の検討を目的とする。以降、本論文では第 2 章に背景埋め込み方式の詳細と問題点を述べ、第 3 章では基本ドットパターンの小タイル化、OCR 技術を用いた提案方式を示す。そして第 4 章で実験方法とその結果、第 5 章で最後に本論文をまとめる。

2. 従来方式

本研究では、従来方式として文献[1][2]を想定する。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ドット群の方向の違いにより「0」「1」のパターンを定義する手法である。図 2 に「0」と「1」の網掛けパターンを示す。文献[1][2]ではプリン

[†] 東京理科大学
Tokyo University of Science

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。この手法は、1 タイル内に 18 ドットを高密度に打ち込むため、背景輝度の増加による画質劣化、ドットパターンが方向性を持つことによる画質劣化が生じ、文書品質の低下を招いてしまう。また、埋込情報量については、一定以上の画素密度が必要なガボールフィルタを抽出に用いるため、密度を一定にした小タイル化は抽出精度の低下を招き、ドット数を一定にした小タイル化は画質劣化を引き起こす可能性があるため、現状の 18×18 画素に 1 ビットが埋込可能な限界情報量と考えられる。さらにガボールフィルタはその計算量の多さで処理速度に対する悪影響も少なからず考えられる。

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単一ドット方式である[3]。概要としては文献[1][2]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し、18×18 画素の中にドットを一つ打ったものを「1」のシンボルパターン、何も打たないものを「0」のシンボルパターンとする。それを図 3 に示す。

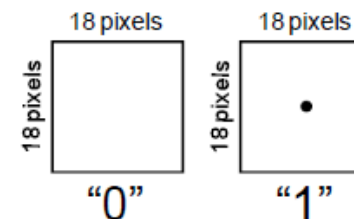


図 3 単一ドットの方式のシンボルパターン

ただ、単一ドットの方式は文献[1][2]の方式と比べてドットの密度が低いためにノイズに対する耐性が低く、印刷時の前提条件として白紙のみを用いることや、折り曲がりや汚れが付着しているものはもちろん、印刷・スキャン時に傾斜があるものは使用できないという問題がある。

また、双方とも文字などの前景があるものに情報を埋め込むと、前景が邪魔になってしまい正確な抽出結果が得られない恐れがあり、実用化に向けての課題点となっている。しかし、文献[1][2]の方式は文字部分と背景の輝度差が少ないため、文字だけを分離すると違和感の生じる文章画像になってしまう一方で、単一ドットの方式はドットの密度が低いと文字部分との濃度差が大きく、文字部分と背景の分離が文献[1][2]の方式と比べて容易であると考えられる。

そのため本研究では、従来手法[1][2]の課題の克服、性能向上のため、

- (1) 視覚的文書品質の向上
- (2) 埋込情報量の向上
- (3) シンプルな抽出演算
- (4) OCR を用いた前景がある印刷物への情報埋め込み

の要件を満たすことを目的としたパターンを定義してその有用性を検証する。

3. 埋め込み情報量向上と OCR を用いた提案方式

3.1 基本ドットパターンの小タイル化

埋め込み情報量を向上させるために、我々はパターンサイズを、標準サイズ 18×18 画素に加え 9×9 画素および 7×7 画素に縮小したパターンを作成し、抽出率を検証した。

図 4 は、600dpi、2160×1296 画素の条件において各パターンサイズで可能な埋込情報量を示している。ドット密度は 7×7 画素のパターンサイズでさえ従来手法より低くなることも分かる。このことは埋込情報量に対する文書品質の劣化が従来手法と比較

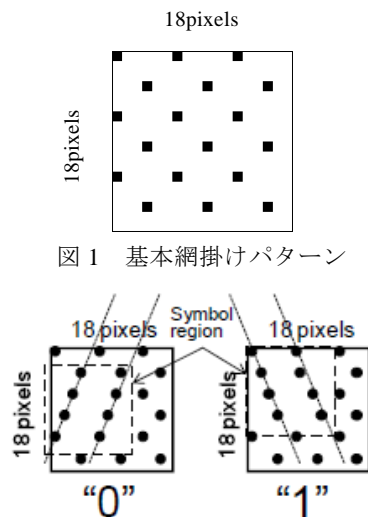


図 1 基本網掛けパターン

図 2 従来手法のドットパターン

して少ないことを表している。

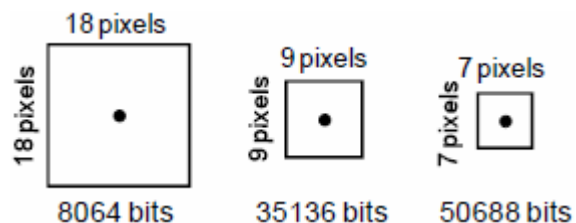


図4 名刺サイズに埋込可能な情報量 (600dpi)

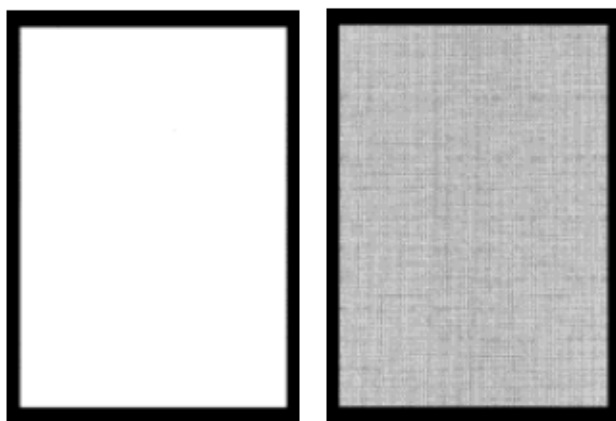


図5 画質比較 (600dpi)

図5は文献[1][2]の方式と単一ドット方式の印刷結果をスキャンした画像である。単一ドットの方式は沖電気の方式と比較して視覚的な違和感が少ないことは明らかで、よりシンプルに情報を埋め込むという目的を果たしたと言える。

以上より単一ドットパターンは、2章で定義した最初の要件、その視覚的文書品質において従来手法と比較しても優れている。これに加え他の要件を満たす以下の長所をも併せ持つ。

(1) パターンサイズは理想的にはドットサイズにまで縮小することが可能であり、埋込み情報量を劇的に増加させる可能性を持つ。

(2) 本手法は情報を1ドットの有無で表わすので、抽出は基本的に加算、比較演算の

組み合わせで可能であり、計算量は従来方式に比較して少ない。

以上より我々は、単一ドットパターン方式を以後の検証に用いることにする。

3.2 OCR を用いた印刷物への情報埋め込み

OCR技術には印刷物を簡単にパソコンで利用可能なデータ(Word形式, Excel形式, PDF形式など)に変換できるOCRソフトが存在する。今回は、OCRソフト(読取革命 Ver.12)を用いて図6の文書画像においての前景部を切り出した後に、情報の埋め込み・抽出を行う。

OCRソフトを用いて図6の文書画像に文字認識を行うと、結果をxml形式として保存することができる。このxml形式は文字の位置情報を持っていて、水平方向における文字の座標や垂直方向における文字の座標を得ることが出来る。この情報を用いて文字部分を切り出し、その部分を避けて前記単一ドット方式によって情報を埋め込んで印刷したものを図7に示す。ちなみに、図6, 7共にA4サイズのを縮小するとドットも識別しにくいので、ドットを識別できる範囲で実験結果として載せる。

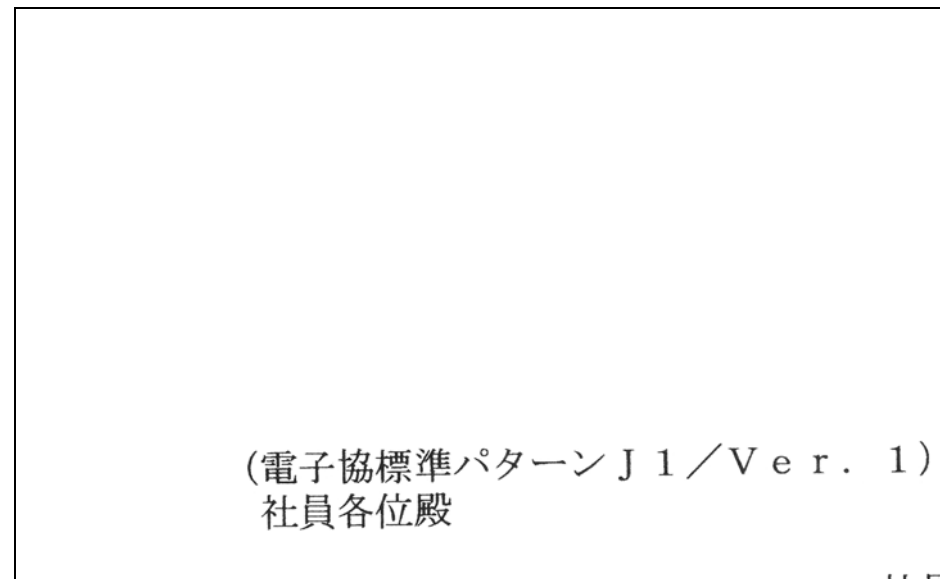


図6 実験で用いた文書画像

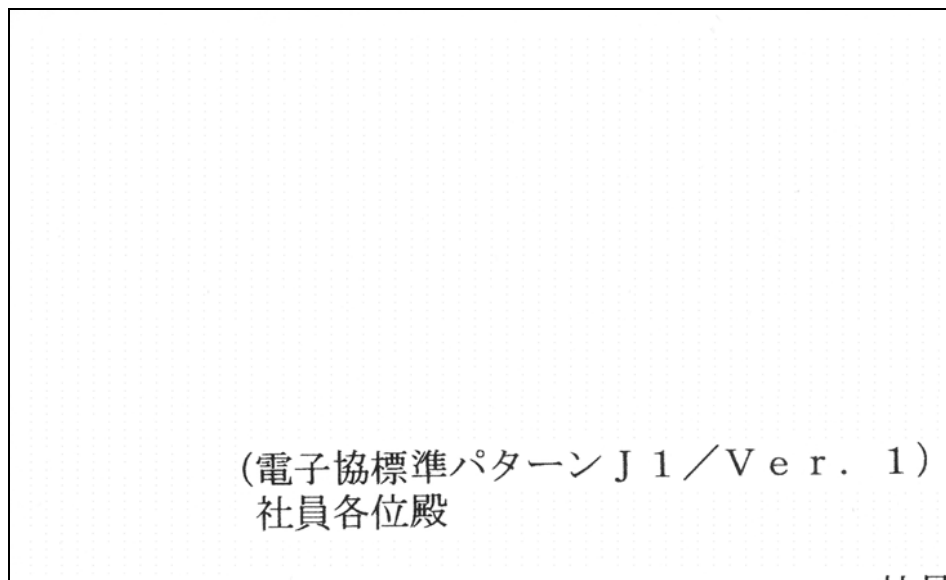


図7 文書画像に前景部を避けて情報「1010」を埋め込んだ画像

4. 実験

4.1 実験・抽出手法

埋め込み情報量を向上させるために、我々はパターンサイズを、標準サイズ 18×18 画素に加え 9×9 画素および 7×7 画素に縮小したパターンを作成し、抽出率を検証した。

抽出動作の処理の流れとしては、ツールを用いて情報を埋め込んだ印刷物（透かし画像）をスキャンして、パソコンに BMP 画像ファイルとして取り込む。取り込んだ BMP 画像に対してツールを用いてドットの検出を行い、透かし画像と抽出した情報を比較し、正しく抽出できたかどうか検証する。

その抽出手法としては、入力された透かし画像の水平、垂直方向の濃度ヒストグラムを計算する。各方向にそれぞれ少なくとも 1 ドットある場合、ヒストグラムにはそれぞれピークが発生する。一定以上の濃度を有するドットが存在する場合、「1」としてドットがあるとし、一定値の濃度を超えない場合は「0」としてドットがないことを抽出結果として示す。

4.2 概要

今回の実験では、文献[1][2]の方式との画質、及び埋込情報量の比較を行うことを目的とするが、印刷紙は白紙、折り曲がり・汚れ無し、印刷・スキャン時の傾斜は無いものとする。本前提条件については別途今後の課題として 6 章で述べたい。

今回の実験条件は以下の通りである。

プリント条件：インクジェット方式、通常文書印刷モード（白黒）、印刷解像度 600dpi

スキャン条件：デフォルト設定、8 ビットグレーモード、読取解像度 600dpi

読取り領域：メモリー消費の観点から名刺サイズ(2016×1296 画素)に限定

5. 結果

次にそれぞれのパターンサイズにおける抽出率を求め、その結果を表 1 に示す。なお、埋込みデータは「1010」を繰り返している。

表 1 各パターンサイズにおける抽出率

	従来方式	単一ドット方式		
		18×18	9×9	7×7
タイルサイズ[画素]	18×18	18×18	9×9	7×7
埋込可能情報量[bit]	8568	8064	32256	50688
抽出失敗情報量[bit]	167	189	682	3828
抽出率[%]	98.05	97.68	97.89	92.45

表 1 より抽出率はどのパターンサイズにおいても 90%以上であることが分かる。これは同期パターン、エラー訂正を使用することにより、現実的な埋込情報量の下で抽出率を 100%にすることが可能であることをも示している。

次に名刺サイズ（2160×1296 画素、600dpi）に情報「1010」を埋め込んだ図 7 の文書画像からドットの抽出を行い、その結果を表 2 に示す。抽出においても埋め込みと同様に文字部分以外からドットの抽出を行う。ちなみに図 7 の前景（文字）部分は名刺サイズのうち 1/8 程度占めている。

この表から名刺サイズと名刺サイズ（前景抜き）は同程度の抽出成功率を得たことがわかる。

表 2 抽出結果の比較

	名刺サイズ	前景抜き
埋込可能情報量[bit]	8064	7450
抽出失敗情報量[bit]	189	108
抽出成功率[%]	97.68	98.55

6. まとめ

本報告では、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[3]を用いて、埋め込み情報量の向上、前景と背景を分離するためにOCR技術を用いて前景部を抽出した後に情報の埋め込み及び抽出を行った。

その結果、従来手法[1][2]の埋込情報量1ビット辺り18×18ドット四方に対し、本手法は7×7四方で6倍強、ドット密度は18×18四方辺り18ドットに対し、7×7四方辺り1ドットで約1/3で、いずれのパターンサイズにおいても90%以上の抽出率を得ることができた。

情報の埋め込みに関して前景部を避けて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することができた。しかし、文字の他にグラフや図などに対しては前景部を避けることができず、印刷スキャン時の紙のずれによって文字認識に微妙な誤りが生じたため、正しく前景を避けて情報を埋め込めない箇所が数ヶ所あった。抽出結果に関して、97%以上の抽出率を得ることができたが、先程と同じ原因で正しく前景部を避けて抽出を行えない箇所があると考えられる。

今後は以上の原因を課題点とし、同期ビット・誤り訂正による抽出率100%を目指し、回転耐性など制約条件の緩和や、ドットの色を人が認識しづらい色に変えることで違和感を少なくし、実用化を目指すための検証実験を進めたい。

参考文献

- [1] 前野蔵人, 須藤正行, “情報漏えい対策に向く印刷文書用電子透かし方式” 信学論 (D), vol.J90-D, no.1, pp.30-39, January 2007
- [2] 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” 信学論 (A), vol.J87-A, no.6, pp.778-786, June 2004
- [3] 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, pp.11-16, July 2008