

印刷物対応電子透かし技術の開発(1)～埋め込み・検出

菅井 豊和

新名 麗

伊藤 浩

鈴木 光義

三菱電機(株) 情報技術総合研究所

1. はじめに

企業機密管理、個人情報保護に対する重要性が増大する昨今、電子データの情報漏洩と併せて、紙媒体を経由した情報漏洩の事故発生件数が増大している。このような情報漏洩への抑止対策の一つとして、印刷物の出所に関する情報を透かしとしてプリンタ出力文書の背景に地紋形式で重畳させることにより埋め込む技術を開発した。万一の情報漏洩時に、漏洩した文書より出所情報を読み取り検出を行うことで漏洩元を特定することが可能となる。

本技術では、地紋の繰り返し配置と多階層化により、文書の部分切り取り耐性と透かしデータ容量拡大の両立を実現した。本稿にて、透かしの埋め込み方式と検出方式について説明する。

2. 処理の流れ

まず、透かしの埋め込みと検出のおおまかな流れについて説明する。印刷者が端末から地紋重畳印刷専用サーバ(プリントサーバ)に文書データの印刷を指示すると、プリントサーバにて透かしに埋め込む出所情報(例えば印刷元・印刷日時・配布先・機密レベル・文書情報等)を取得し、これらの情報より生成した地紋パターンを印刷スプールデータへ重畳してプリンタ経由で文書を出力する。文書漏洩が発生した場合、漏洩した文書紙面をスキャナで読み取り、ソフトウェアによる透かし情報検出処理を実行する。

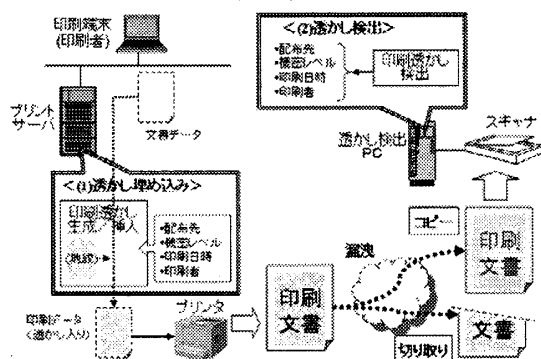


図 1 透かしの埋め込みと検出の流れ

Development of Watermarking Method for Printed Documents (1)~ Embedding & Detecting method

Toyokazu Sugai, Urara Shimmyo, Hiroshi Ito, Mitsuyoshi Suzuki, Information Technology R&D Center, Mitsubishi Electric Corp.

3. 地紋パターン生成方式

透かし埋め込みの際の地紋パターンの生成について説明する。まず、透かし検出時のビット誤り補正を目的として、透かしとして埋め込む情報に対して誤り訂正冗長符号化を行う。本方式では、リードソロモン符号 RS(255, 239)を利用して冗長符号化を行っている。

次に、冗長符号化ビット列を地紋パターンに変換する。正方形のピクセルパターン(以後、「基本ブロック」と呼ぶ)を基本構成単位とし、これを A4 紙面全面に縦横に並べて地紋パターンを作成する。ビット 0/1 を表現するパターン(基本ブロック 0/基本ブロック 1)をそれぞれ定義し、基本ブロック毎に 1 ビットの情報を割り当てる。また他に、透かし検出時に印刷やスキャンの際に発生する歪みを補正する基準座標を示す「補正基準点パターン」を設定する。図 2 に各パターンの一例を示す。



左: 基本ブロック 0
中央: 基本ブロック 1
右: 補正基準点パターン

図 2 地紋パターンの一例

文書の部分切り取りに対して検出耐性を持たせるため、同一の地紋パターンを紙面に繰り返し配置する。例えば、縦 4 分割、横 4 分割した 16 の分割領域に対して、各々同一地紋パターンを配置することで、最小でも縦 1/4、横 1/4 の矩形領域より埋め込みビット列が復元できる。さらに、繰り返し数を変えた複数の地紋パターンを重畳させ、データを多階層化させることにより、検出耐性の異なる複数の透かしデータを共存させることが可能となる。これらの工夫により、重要なデータには部分切り取り耐性を保ちつつ、データ全体の容量を拡大することが可能な方式となっている。

4. 検出方式

透かしの検出方式について述べる。まず、印刷文書より取得したスキャン画像データの全領域を対象に補正基準点パターン画像との相関(正

規化相互相関)を計算し、相関の高い部分を補正基準点の座標として取得する。

次に、文書の印刷時やスキャン時に発生する微小な拡大縮小、回転、平行移動、スキュー等の歪みを補正するため、印刷文書のスキャナ読み取りで取得したスキャン画像データに対してアフィン変換による画像補正を行う。取得した補正基準点の座標と、元の地紋画像の補正基準点位置との対応関係をハフ変換[1]を利用して特定し、最終的にアフィン変換パラメータを決定して画像補正を実行する。

補正後のスキャン画像では、基本ブロックが縦横に整列していることを前提にして良いため、画像内のブロック領域毎に基本ブロック 0 と基本ブロック 1 の相関をそれぞれ計算していき、より相関値の高い方を対応するビットと判定することでビット抽出を行っていく。

次に、スキャン画像全体より抽出した検出ビット列に対し、RS (255, 239) 符号の誤り訂正を行う。さらに、繰り返し埋め込みが行われている領域同士の検出ビット平均をとることで、より精度の高い誤り訂正が可能となる。(ただし、部分切り取り文書からの検出の場合は、複数の繰り返し埋め込み領域が残存することは保証されないため、必ずしもビット平均処理が実行できるとは限らない。)

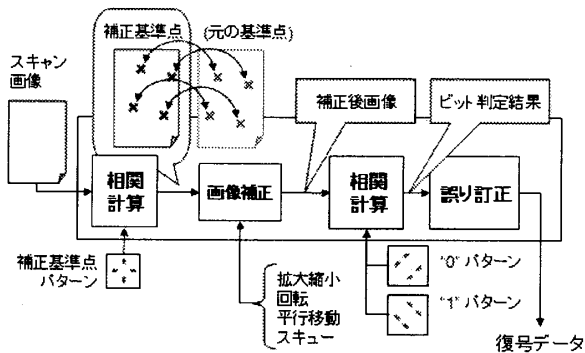


図 3 透かしの検出方式

5. 実現性能

A4 サイズ文書に対して、40/216/256/512 バイト(計 1024 バイト)の 4 種類のデータを多階層化して埋め込みを行った。それぞれ 64/16/4/1 回繰り返し埋め込みを行ったため、40 バイトデータが最も強い検出耐性を備えていることになる。

全体的な印字密度が 10%程度の標準的な A4 文書データを用いて、地紋入り印刷文書より部分切り取りを行った画像に対して、本透かし検出方式で検出処理を行った結果を示す。

図 4 は、解像度 600dpi でスキャンした画像(約 5000×7000 ピクセル)よりおよそ縦 1/4、縦

1/4 サイズ矩形相当の 1300×1800 ピクセルの部分領域への検出処理結果として、各部分画像内の印字密度と検出バイト数との関係をプロットしたグラフである。

印字密度が比較的低い(8%以下)部分では 100%の割合で 256 バイト検出に成功しているのに対し、印字密度 8%を超えると 40 バイト検出にとどまっているケースが出現し、全体としては約 9 割超のケースで 256 バイト検出に成功した。

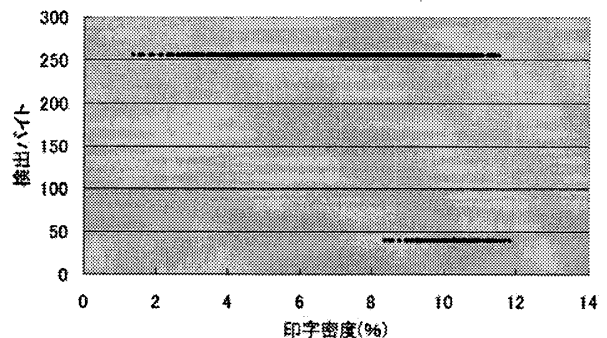


図 4 部分切り取り画像からの検出結果

また、図 5 は複写コピーした文書についての検出結果を示しているが、この場合は若干印字密度が低い場合でも 40 バイト検出にとどまっているケースが見られ、256 バイト検出成功率は約 7 割程度となっている。

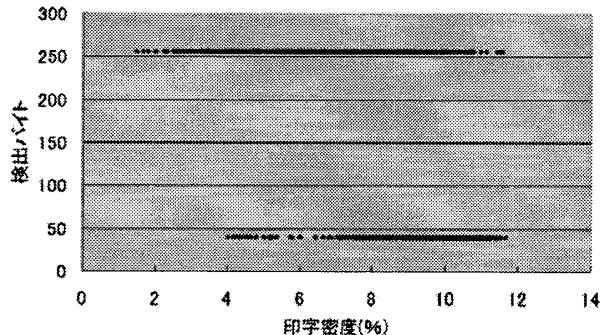


図 5 複写コピーからの部分切り取り検出結果

なお、図 4・図 5 いずれの場合も、40 バイト以上の検出には 100%成功している。

6. まとめ

紙媒体への情報漏洩抑止対策として、プリンタ出力文書の背景に地紋形式で出所情報を重畳させることにより埋め込む技術を開発した。

参考文献

[1] D. H. Ballard, "Generalizing the Hough Transform to Detect Arbitrary Shapes," Pattern Recognition, vol. 13, no.2, pp. 111-122, 1981