

カラー難視性パターンを用いた情報付加手法に

おける前景耐性改善の提案

Proposal of the Robustness by using the Improved Color Character in Information Hiding based on the Artificial Fiber Pattern

金田 北洋†
Kitahiro Kaneda

鬼頭 祐太†
Yuta Kito

岩村 恵市†
Keiichi Iwamura

1. まえがき

近年の情報漏洩の傾向は、Winnyに代表される個人の過失割合が高い事象から、ウィキリークスに代表される組織的な情報暴露に変化しつつあり、悪質化してきていると言えよう。このような背景から、企業、公的機関等の情報漏洩防止に対する意識は以前にも増して高まってきている。一方事実として、最もインシデント件数が多い漏洩媒体・経路は、印刷物で、2009年度は全体の73%となっておりその割合は年々増加している[1]。また、ネット上の情報暴露も元をたどると誰でも専門知識が無くてもアクセスできる紙文書であることが多い。そのため抑止技術として印刷物への情報埋込み・抽出技術の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文字や画像などのコンテンツ自体に電子透かしを埋め込む手法(不可視)、下地(例えば、紙そのもの)にバーコードや特殊なパターン等を埋め込む手法(可視)に分類される。これらはそれぞれ長所、短所があり、攻撃耐性・埋込み情報量・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。

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、幾何変換に対する脆弱性を解決するために紙メディア自体をコンテンツとした難視性パターンを考案し、それを利用した印刷文書に対する難視性パターンを用いた新しい情報付加手法を提案している。難視性パターンとは、情報埋め込みのためのテクスチャパターンのことで、紙繊維の凸凹の特徴を活かしているため、視覚的な違和感を感じにくいと考えられる。文献[2]では、難視性パターンの原理について詳細に説明し、文字等の前景が無い状態に対して情報埋込み・抽出の実験を行い、印刷や回転等の基本的な耐性があることを証明した。また文献[3]では、難視性パターンに後から加えられる手書きメモやセロテープ貼付け等の攻撃に対する耐性の実験を行い、問題の無いことを確認している。また、前景を想定して、難視性パターンをマゼンタ濃度80%(グレー濃度換算16%)で着色された6種類のエッジパターンに重畳し、それを無着色の難視性パターンと重ね合わせても情報を読み取ることが可能であることを確認した。

難視性パターンは印刷文書に使用されることを想定しているので、文字に代表される前景構造物がその上に印

刷された場合の耐性が無ければ、実用的には使えない。これまでの研究では、実際の文字を印刷した場合の検証までは行われていない。文献[3]で難視性パターンを重畳させたマゼンタを使用した理由は、ベタ塗り前景の場合、ベタ塗り部分のDC成分と、前景と背景の濃度差が高いため、そのエッジのリング成分が難視性パターンを識別するために特徴付ける周波数領域に漏れ込み、0, 1に相当する難視性パターンの識別を困難にしてしまうことがあげられる。そのため、背景との濃度差が小さいグレー濃度でベタ塗りではない難視性パターンを重畳した文字を構成する基本パターンを作成し、そこから実験をスタートさせた。マゼンタを使用した理由は、カラーで識別しやすいことと、プリンタの原理上、純色で濃度を調整できることにある。文献[4]は、実際の文字にマゼンタとシアンの難視性パターンを重畳させた場合の抽出方法の改良、輝度調整による難視性パターン画像の正規化、について述べられているが、そもそも従来手法における前景と背景の濃度差の問題点が明らかにされていない、濃度差と共に難視性パターンも取り除いてしまう可能性がありその効果は限定的である、という課題が残されている。

本研究の目的は、文字を前景とした場合の耐性を実験的に検証することにある。そのため、本論文では、まず前景としての文字を背景の難視性パターンと濃度差を付けた難視性パターン(以後カラー難視性パターン)を用いることによる問題点の把握を行う。その上で、濃度差の影響を最小限にするためのより効果的な抽出改善手法を提案し、情報埋込み・抽出実験によりその妥当性を検証する。

2. 従来手法

2.1 難視性パターン

現状の印刷物への情報埋込み・抽出技術は大別すると電子透かしを埋め込む手法(不可視)、下地にバーコードや特殊なパターン等を埋め込む手法(可視)の2つに分類される。文字、あるいは図形といったコンテンツに情報を埋め込む電子透かし方式は、文書品質の劣化は小さいが、埋込み容量や幾何変換耐性などに課題がある。また、下地に埋め込む手法は埋込情報量が比較的多くコンテンツに左右されず安定しているが、線パターン、ドットパターンによる視覚的な文書品質が劣化する。このような従来方式の課題を解決するために、印刷文書の下地に情報を埋め込む手法と紙メディア自体をコンテンツと考えた電子透かしの方法を融合させた「難視性パターン」を用いた新しい情報埋込手法がある[2]。

†東京理科大学工学部, 東京都千代田区九段北 1-14-6
Tokyo University of Science, 1-14-6 Kudan-Kita, Chiyoda-ku,
Tokyo, 102-0073 Japan

難視性パターンは、紙をミクロ的に解析すると見える紙の繊維の凸凹の特長を活かしている。図1はそれぞれ、業務用高現像度スキャナを使用して紙を2400dpiでスキャンした画像と、その周波数分布を表したものである。これにより紙自体はそれほど高い空間周波数を持ってないことがわかる。難視性パターンを利用することで、低視認性による高画質化、印刷耐性、幾何変形・紙形状変形耐性といった効果が期待できる。

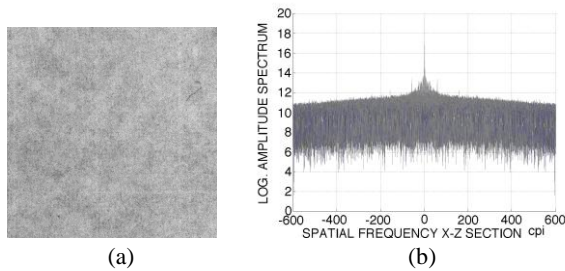


図1 高現像度スキャンによる紙メディアの特性

2.2 難視性パターンの生成手順

図2と共に以下に難視性パターンの生成手順を示す。詳細は文献[2]を参照。

(1) ターゲットブランク紙をグレースケールでスキャンし画像化し、0.5inch四方のパターンを切り取る。これをオリジナルパターン $o(x,y)$ とする。

(2) オリジナルパターン $o(x,y)$ に対し、ハイライト、シャドウ、ガンマのバランスをPhotoshopの自動レベル補正で最適化し、紙繊維の凹凸を自然に観察できるようにする。この結果を以後基準パターン $s(x,y)$ とする。

(3) 基準パターン $s(x,y)$ を二次元フーリエ変換し、結果を $S(u,v)$ とする。また、その直流成分を S_{dc} とする。

(4) $S(u,v)$ の低周波成分を円状にくり抜く、以後このカットオフ周波数を r cpi (cycle per inch), 結果を $S'(u,v)$ とする。ここで、 r は情報 (1, 0) に応じて r_1 および r_0 をとるものとする。残された $S(u,v)$ の交流成分を $S_{ac}(u,v) (= S(u,v) - S_{dc})$ とする。

(5) S_{dc} を戻し、さらに難視性(濃度)を調整する難視性調整係数 α (0~1) を $S_{ac}(u,v)$ に乗じる。この処理結果を $S''(u,v)$ とする。

(6) $S''(u,v)$ を二次元逆フーリエ変換し、難視性パターン $p(x,y)$ を求める。実際には(4)、(5)のプロセスよりカットオフ周波数 r_1 および r_0 に対応した $S'(u,v)$ 、 $S''(u,v)$ が生成され、難視性パターンもそれに対応した $pr_1(x,y)$ 、 $pr_0(x,y)$ が求められる。

(7) (6)までのプロセスで得られた基準パターン $s(x,y)$ と難視性パターン $pr_1(x,y)$ 、 $pr_0(x,y)$ を以後当該用紙の情報埋込用パターンセットとし、情報に応じてターゲット用紙に印刷する。

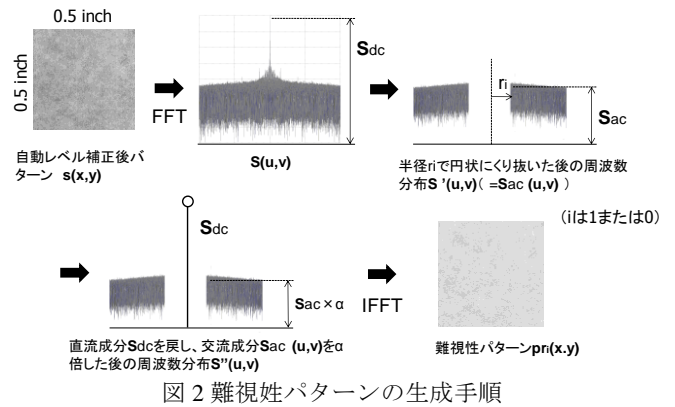


図2 難視性パターンの生成手順

2.3 難視性パターンの従来手法の抽出手順

以下難視性パターンの従来手法の抽出手順を示す。なお生成の際に作成した基準パターン $s(x,y)$ を用いる。

(1) 難視性パターン $p(x,y)$ (情報埋込み用、実際には $pr_1(x,y)$ 、 $pr_0(x,y)$) が印刷された紙をグレースケールでスキャンし、画像化し、0.5inch四方に分割する。

(2) (1)で得られたパターンに対し自動レベル補正を施し、この結果を $p'(x,y)$ (実際には $pr'_1(x,y)$ 、 $pr'_0(x,y)$) とする。

(3) $p'(x,y)$ に二次元フーリエ変換を施し、その振幅スペクトルを $P'(u,v)$ とする。

(4) 基準パターン $s(x,y)$ に二次元フーリエ変換を施し、その振幅スペクトルを $S'(u,v)$ とする。

(5) $P'(u,v)$ の $S'(u,v)$ に対する低周波成分減少度合いを定量的に評価するために強度比 $d(R)$ を式(1)のように定義する。

$$d(R) = \frac{\iint_0^R |P'_{ri}(u,v)| dudv}{\iint_0^R |S'(u,v)| dudv} \quad (1)$$

ここで、 i は1または0。

(6) 予め強度比に関連付けられたビット情報を埋込み情報として抽出する。

3. カラー難視性パターンによる文字前景耐性の確認

3.1 実験概要

文献[3]では、6種類のエッジパターンをグレー濃度換算16%のマゼンタ濃度80%で表現した前景重畳難視性パターンから情報抽出が可能であることを確認した。ここではまず、前景として実際の文字を背景の難視性パターンと濃度差を付けた難視性パターン、カラー難視性パターンとして生成し、濃度差の大きさが文字前景耐性に与える影響を実験的に確認する。

3.2 使用機材

今回の実験で使用した機材は以下である。

- (1) プリンタ
 - ・ Canon 製 MP970 (印刷解像度 9600×2400dpi)
- (2) スキャナ
 - ・ Kodak 製 IQSmart2 (最大光学解像度 4300dpi)

(3) 難視性/基準パターンサイズ

- ・基準パターン (0.5inch 四方, 1200×1200 ピクセル)
- ・難視性パターン (1inch 四方, 2400×2400 ピクセル)

(4) 印刷条件

- ・カラー, 高画質, 2400dpi

(5) 読取り条件

- ・グレースケール (8bit), デフォルト設定, 2400dpi

(6) 紙メディア

- ・ビジネスカットペーパー

(7) 使用コンピュータ, 及びソフト

- ・PC: HP xw8600Workstation Intel(R) Xeon(R) CPU E5420 @ 2.50GHz 3.25GB RAM
- ・MATLAB Ver7.5.0(R2007b)
- ・Photoshop cs2 Ver9.0

3.3 基本パラメータ

基本パラメータであるカットオフ周波数 r_i , 難視性調整係数 α , 及び積分半径 R の設定は文献[2]~[3]で使用している $R=10$, $\alpha=1/2$, $r_1=2$, $r_0=20$ とする。

3.4 実験方法

今回の実験はリファレンス機材として Canon 製プリンタ MP970, Kodak 製高解像度スキャナを用い, ビジネスカット紙に対して行った。3.3 の基本パラメータを使用した従来の難視性パターン上に,

マゼンタ濃度 80% (グレー濃度換算 16%, 以下 M80%)

マゼンタ濃度 100% (グレー濃度換算 20%, 以下 M100%)

シアン濃度 100% (グレー濃度換算 22%, 以下 C100%)

のカラー難視性パターンで構成される 0 から 9 までの数字を重畳させて印刷し, 抽出が可能か否かの実験を行った。文字サイズは, 2400dpi 換算で 400pt とした。これは今回のパターンサイズ 0.5inch 四方に数字がすべて入る程度の目安で設定した。以下に実験手順を述べる。

(1) A4 ビジネスカット紙をベースに Kodak 製 IQSmart2 を使用し 2.2 の手順で基準パターン $s(x,y)$ と難視性パターン $pr_1(x,y)$, $pr_0(x,y)$ を作成する。図 3, 4 にそれらを示す。この際 3.3 の基本パラメータを使用する。

図 3 基準パターン $s(x,y)$ 図 4 難視性パターン $pr_1(x,y)$, $pr_0(x,y)$

(2) 次に Adobe Photoshop cs2 を使用して難視性パターン $pr_1(x,y)$, $pr_0(x,y)$ 上に, それぞれ M80%, M100%, C100% の数字のカラー難視性パターンを重ね合わせる。それらを図 5~7 に示す。

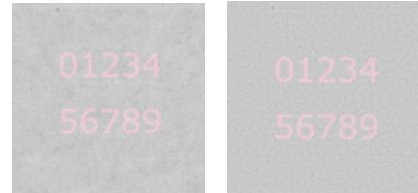


図 5 M80% カラー難視性パターン

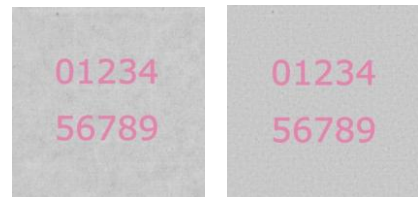


図 6 M100% カラー難視性パターン



図 7 C100% カラー難視性パターン

(3) カラー難視性パターンを図 8 に示すように $pr_1(x,y)$, $pr_0(x,y)$, $pr_1(x,y)$ の順に横に 3, 縦に 7 の計 21 個 ($pr_1(x,y)$ が 11 個, $pr_0(x,y)$ が 10 個) を配置する。これをビジネスカット紙に印刷する。

pr1(x,y) データ 番号1	pr0(x,y) データ 番号1	pr1(x,y) データ 番号2
pr0(x,y) データ 番号2	pr1(x,y) データ 番号3	pr0(x,y) データ 番号3
pr1(x,y) データ 番号4	pr0(x,y) データ 番号4	pr1(x,y) データ 番号5
pr0(x,y) データ 番号5	pr1(x,y) データ 番号6	pr0(x,y) データ 番号6
pr1(x,y) データ 番号7	pr0(x,y) データ 番号7	pr1(x,y) データ 番号8
pr0(x,y) データ 番号8	pr1(x,y) データ 番号9	pr0(x,y) データ 番号9
pr1(x,y) データ 番号10	pr0(x,y) データ 番号10	pr1(x,y) データ 番号11

図8 難視性パターン読取り実験シート

(4) Kodak製 IQSmart2 を使って、各難視性パターンが埋め込まれている位置を 1inch 四方のサイズにて 8bit グレースケールでスキャンしていく。その後は (1) で生成した基準パターン $s(x,y)$ を用いて、2.3 の手順で 21 個の強度比を算出する。

3.5 結果

M80%, M100%, C100% の強度比分布の結果を前景無しとのそれと並べ、それぞれ図9~11に示す。横軸のデータ番号は、図8のデータ番号に対応している。

これらの結果より、M80%では、現状の抽出手法でも $pr1(x,y)$, $pr0(x,y)$ では強度比に十分な差があるため、カラー難視性パターンにおいてもその識別は可能である。さらに、前景無しの強度比分布と共通の閾値で識別することが可能であることも分かる。M100%の強度比分布はM80%と比較して十分ではないものの識別可能な差は認められるが、前景無しの強度比分布と比較すると全体的に高い値となっており、共通の閾値で識別することは困難である。C100%になると、強度比の差も、前景無しとの共通性も無く、 $pr1(x,y)$, $pr0(x,y)$ の識別は不可能であることが分かる。

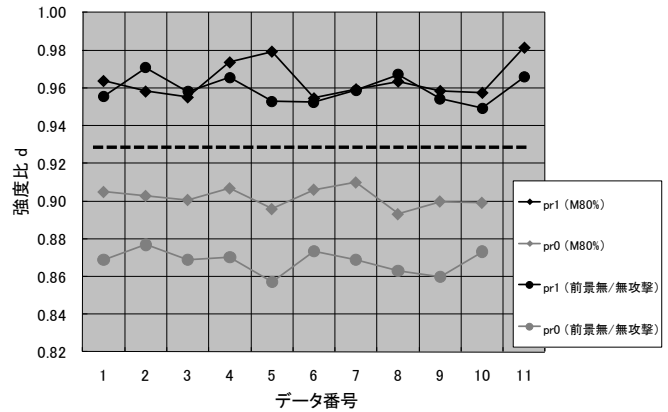


図9 M80%の強度比分布

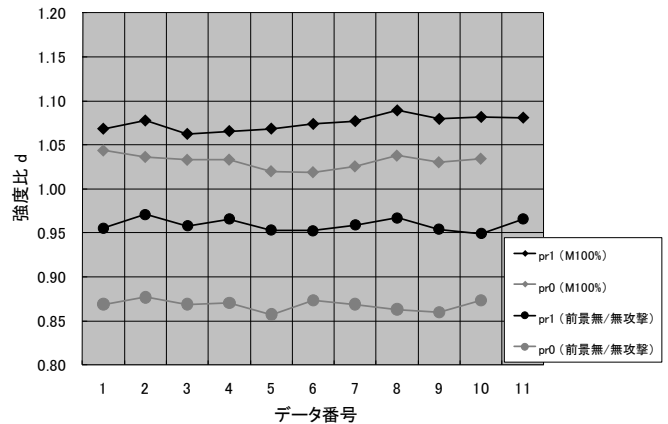


図10 M100%の強度比分布

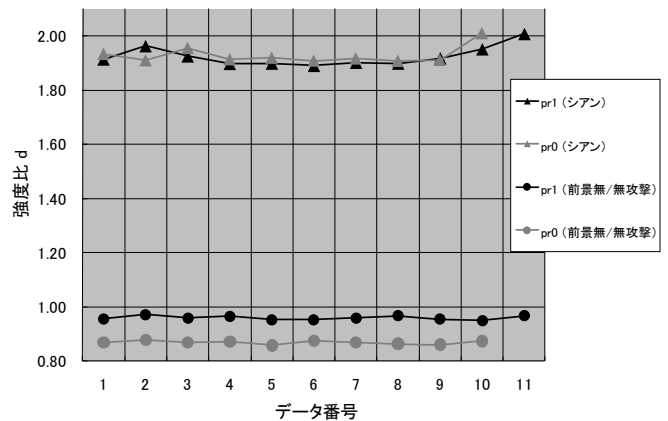


図11 C100%の強度比分布

4. 前景の影響を抑えるための提案

4.1 提案手法概要

これまでの実験結果より、カラー難視性パターンを利用した前景文字耐性は、M80%, すなわち、グレイ濃度換算 16%は問題無いが、それより濃度差の大きいグレイ濃度 20%, 22%の場合は、現状の抽出手法では情報識別が困難であることが判明した。

そのため、ここでは、積極的に濃度変化を除去するため、これまでグレースケールで読み取っていた難視性パターンをカラーで読み取り、前景に該当する色成分を特定し、その影響を除去する手法を提案する。難視性パタ

ーンの生成手順については従来通りで問題無いと思われるため、以下変更した抽出手順のみを説明する。

4.2 提案手法による抽出手順

以下難視性パターン今回の提案手法による抽出手順を示す。なお生成の際に作成した基準パターン s はそのまま用いる。

(1) 難視性パターン $p(x,y)$ (情報埋込み用、実際には $pr1(x,y)$, $pr0(x,y)$) が印刷された紙を RGB カラーでスキャンし、画像化し、0.5inch 四方に分割する。

(2) 分割した画像を CMYK に変換し、シアン、マゼンタ、イエロー成分の画像を抽出する。

(3) 閾値 c , m , y を定め、取り出したシアンの閾値 c 以上、マゼンタの閾値 m 以上、イエローの閾値 y 以上の成分をそれぞれカットした画像を作成する (c , m , y の範囲は 0~1)。閾値 c , m , y を設定した意味については、5.考察で説明する。

(4) 作成した画像をグレースケールに変換し得られた画像に対し自動レベル補正を施し、この結果を $p'(x,y)$ とする。

(5) 2.3 (3) ~ (6) のプロセスを実行し、強度比を算出して情報を抽出する。

以下実験により、今回提案した手法の妥当性を検証する。

4.3 使用機材、基本パラメータ

使用機材は、読み取り条件に、RGB (24bit, 8bit×3) を追加する以外は 3.2 と同様、基本パラメータは、3.3 と同様とする。

4.4 実験方法

3.4 (1) ~ (3) までは共通。ただし、今回実験に使用したカラー難視性パターンは、M100%、C100%、及び C100%+M100%の混色を新たに追加した。図 12 にそれを示す。

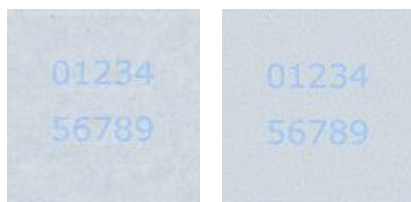


図 12 C100%+M100%カラー難視性パターン (0-9)

抽出手順は、Kodak 製 IQSmart2 を使って、各難視性パターンが埋め込まれている位置を 1inch 四方で 24bit(8bit×3)RGB カラーでスキャンしていく。その後は 3.4 (1) で生成した基準パターン $s(x,y)$ を用いて、4.2 の手順で 21 個の強度比を算出する。

4.5 結果

- (1) C100%の前景で 4.2(3)で定める閾値は、
 $c=0.1(10\%)$
 $m=0$
 $y=0$

として抽出を行った。その結果を図 13 に示す。横軸のデータ番号は、図 8 のデータ番号に対応している。この結果より $pr1(x,y)$, $pr0(x,y)$ では強度比に差が出ているため、C100%の文字を前景としたカラー難視性パターンにおいてもその識別が可能、すなわち情報抽出が可能であることが分かる。

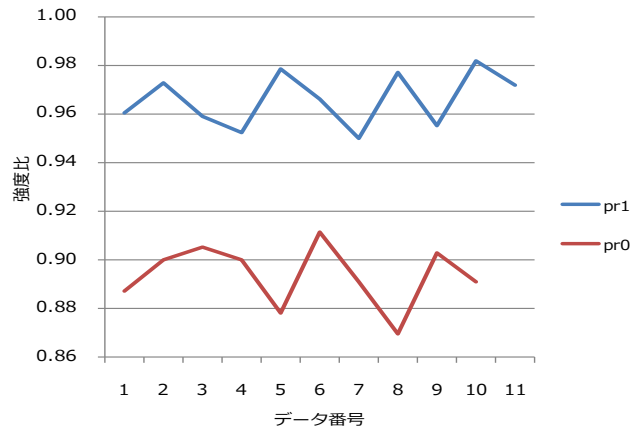


図 13 C100%の強度比分布

- (2) M100%の前景で 4.2(3)で定める閾値は、
 $c=0$
 $m=0.1(10\%)$
 $y=0$

として抽出を行った。その結果を図 14 に示す。この結果より、M100%の文字を前景としたカラー難視性パターンでも、難視性パターンによる情報抽出が可能であることが分かる。

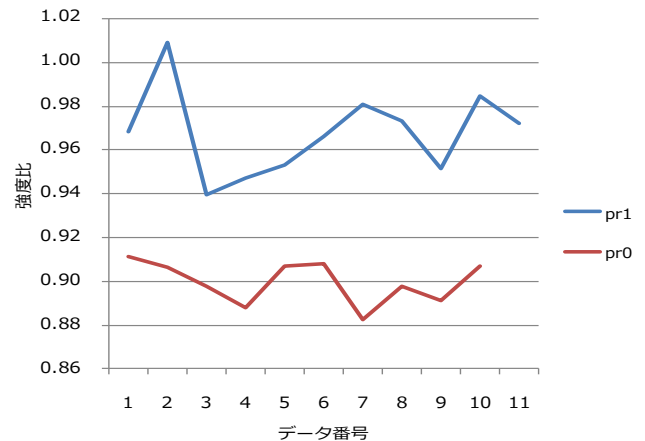


図 14 M100%の強度比分布

- (3) C100%+M100%の前景で 4.2(3)で定める閾値は、
 $c=0.1(10\%)$
 $m=0.1(10\%)$
 $y=0$

として抽出を行った。その結果を図 15 に示す。この結果より、C100%+M100%の文字を前景としたカラー難視性パターンでも、難視性パターンによる情報抽出が可能であることが分かる。

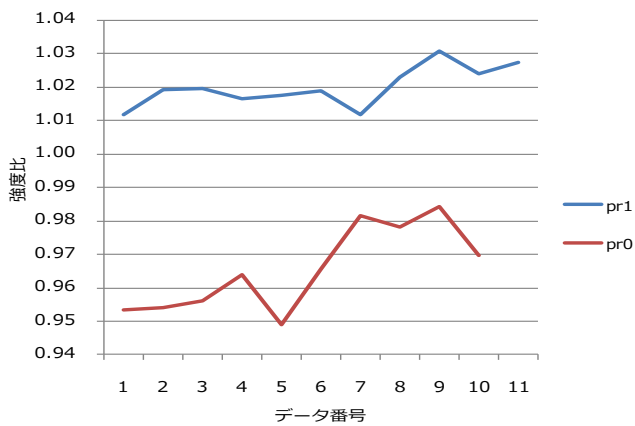


図 15 C100%+M100%の強度比分布

(4) 前景無しの場合でも本提案手法が使用可能であるか否かの実験も行った. 図 4 の難視性パターン $pr1(x,y)$, $pr0(x,y)$ を用いて, RGB カラーで読み取り, グレースケールに変換して抽出を行っている. その結果を図 16 に示す. これより前景無しの場合でも難視性パターン $pr1(x,y)$, $pr0(x,y)$ の識別が可能であることが分かる.

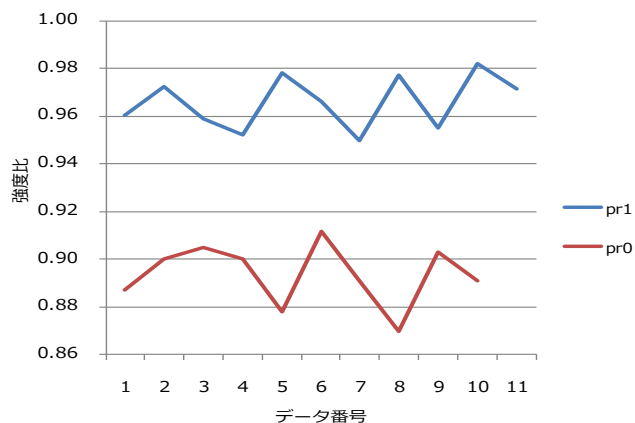


図 16 前景無しの強度比分布

5. 考察

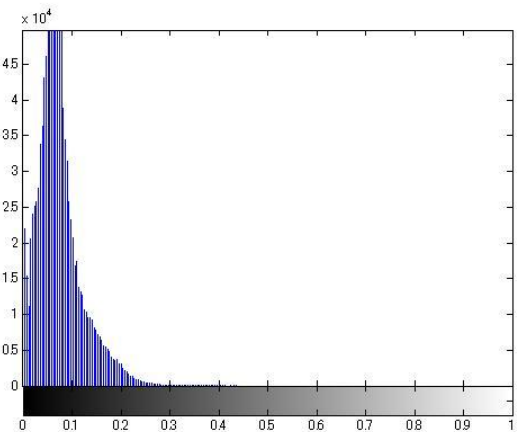
今回の提案手法のポイントは, カラーで読み取りを行い, 前景を削除することにある. 具体的に説明するために, 前景なしの難視性パターン $pr1(x,y)$ (図 4(a)) と, C100%の数字で色付けしたカラー難視性パターン $pr1(x,y)$ (図 7(a)) の 0.5inch 四方のシアン成分のヒストグラムをそれぞれ図 17, 18 に示す. 横軸が輝度 (8 ビットを 0~1 で正規化している), 縦軸が累積ピクセル数となっている. 比較すると, 横軸がほぼ 0.1 以上でカラー難視性パターンの方がシアンの成分が多く出ている. よって 0.1 以上の部分がシアンの前景と推測することができるので, 閾値 c を $c=0.1$ とし, それ以上をカットすることで強度比による識別が可能になったと考えられる. 実際に, C100% カラー難視性パターンのデータ番号 1 の $pr1(x,y)$ をスキャンし 0.5inch 四方に分割した画像と, 閾値 $c=0.1$ とし C をカットした画像と, それをグレースケールに変換した画像を図 19 に示す. このプロセスより視覚的に前景の影響

が除去されていることが明らかで, それは図 11 と 13 の強度比分布の違いにも現れている. これはマゼンタでも同様である.

しかしながら, イエロー成分に関しては, 反射率が紙本来の反射率に近い為, それらの輝度分布の差分が小さく, スキャンする際イエロー成分として分離されにくいという問題がある. 実際に今回の予備実験で観察した結果を図 20, 21 に示す. 図 20 は, 難視性パターン $pr1(x,y)$ の前景として Y100%の数字で色付けした画像, 図 21 は, その 0.5inch 四方の画像のイエロー成分のヒストグラムである. 図 21 より, イエロー成分のヒストグラム自体が小さいため, 少なくともシアン, マゼンタと同様の尺度で下地と前景を分離することが困難であることが分かる. しかしながら, 実際の印刷文書においては黒 (C100%+M100%+Y100%) をはじめとした CMY の混色が多用されるため, この問題は避けては通れない.

これに加え, c , m , y の閾値を前景色に応じて可変させる必要もあるが, その対策としては, 4.2(2)の後に前景色の色成分の割合から判定を行い, それに応じて閾値を決定するなどの処理が必要であると考えられる. イエロー成分がある場合に関しても, 前景色の色成分の割合から判定を行い, 補正するようなことを行わなければならない.

また, 前景無しの強度比分布図 16 と, C100%, M100%のそれである図 13, 14 を比較するとそれらは共通の閾値で $pr1(x,y)$, $pr0(x,y)$ の識別が可能であることが分かるが, C100%+M100%の強度比分布図 15 と比較すると, 全体的に高い値を示し, 共通の閾値で識別することが困難であることが分かる. この現象から, 混色の場合は単純に c , m , y の閾値を足し算で決めて該当成分をカットしても前景の濃度差が強度比の分布を崩すほどでは無いが, 影響を及ぼす程度に残っていると推測することができる.

図 17 前景無し $pr1(x,y)$ の C 成分のヒストグラム

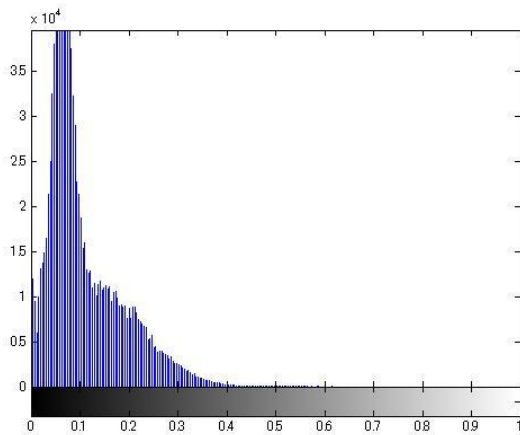
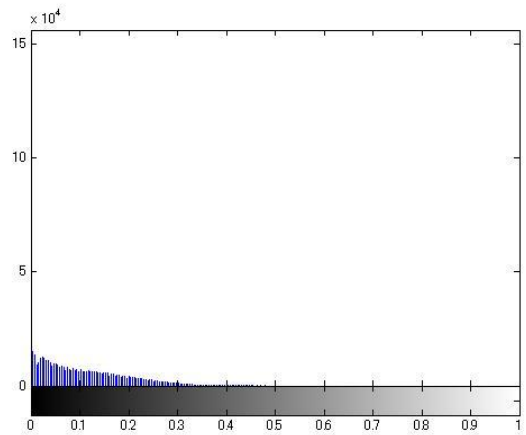
図 18 C100% $pr1(x,y)$ の C 成分のヒストグラム

図 21 Y 成分のヒストグラム



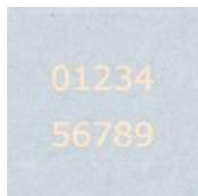
(a) シアン成分除去前



(b) シアン成分除去後



(c) グレースケール変換後

図 19 C100% $pr1(x,y)$ の前景を除去する過程図 20 Y100% カラー難視性パターン $pr1(x,y)$

6. まとめ

本論文では、濃度差のある難視性パターンとしてこれまでのエッジパターンによる耐性検証の結果を元に、まず前景として文字に背景の難視性パターンと濃度差を付けた難視性パターン、カラー難視性パターンとして定義し、その前景耐性の確認を行った。その結果、グレー濃度換算 16%の数字に対する耐性を確認することができた。しかしながら、グレー濃度換算 20%、22%のカラー難視性パターンによる数字の前景耐性は低下し、従来の抽出手法で難視性パターン $pr1(x,y)$ 、 $pr0(x,y)$ を識別することが困難であることが分かった。

次に、濃度差の大きい前景の影響を最小限にするための抽出手法として、カラーでスキャンし、そのヒストグラムから前景に該当する色成分のみを除去する手法を提案し、抽出実験によりその妥当性の検証を行った。結果として、識別困難であった M100% (グレー濃度換算 20%)、C100% (グレー濃度換算 22%)、またその組合わせならば識別が可能であることを実証した。

今後は、イエロー、及び混色のカラー難視性パターンの読み取りを可能とする手法の改良を進め、実用化を目指したい。

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, 平野 謙二, 藤井 雄騎, 岩村 恵市, 半谷 精一郎, “難視性パターンを用いた印刷文書に対する情報付加手法の提案”, 情報処理学会論文誌, Vol.50, No.9 (2009).
- [3]金田 北洋, 藤井 雄騎, 鬼頭 祐太, 岩村 恵市, 半谷 精一郎, “難視性パターンを用いた情報付加手法における攻撃耐性の改善”, 電子情報通信学会論文誌, Vol.J93-A, No.2 (2010).
- [4]金田 北洋, 鬼頭 祐太, 岩村 恵市, “難視性パターンを用いた情報付加手法における前景耐性改善の提案”, 情報処理学会研究報告 CSEC 47(7) pp.1-6 (2009-12).