

H-001

印刷文書の改ざん検出に適用可能な特徴抽出方式に関する考察

A Consideration of the Feature Extracting Method for Tamper Detection in Printed Documents

高橋 由泰
Yoshiyasu Takahashi山田 隆亮
Takaaki Yamada洲崎 誠一
Seiichi Susaki

1. まえがき

2000年以來、日本政府はIT戦略を進め、情報インフラ整備という面からは、世界最高の水準に到達したと言われている[1]。しかしながら、全ての情報インフラが必ずしも利用者の利便性に結びついているとは言えず、政府のIT新改革戦略評価専門調査会において、さまざまな分野の現状分析、課題、解決の方向性の提言が行われている[2]。

[2]では、電子行政分野において、「フロントオフィス改革とバックオフィス改革の連携強化」における解決の方向性として、「行政機関と民間との連携の推進」を図る上で、「コンビニエンスストア等民間サービス事業者を活用した行政手続等に関する『電子交付サービス』の実現可能性について検討すべき」とされている。

ここで言う電子交付サービスとは、遠隔地における印刷によって、電子申請した証明書類の交付を受けるサービス（遠隔地印刷型）である。なお、電子的な証明書データを、電子的に送信されることによって交付を受けるサービス（完全電子型）も考えられるが、この場合電子的な証明書データの提出先が、電子的な証明書データに対応する必要があるため、より普及が難しいと考えている。本稿では、遠隔地印刷型に絞って考察する。

遠隔地印刷型の電子交付の一形態としては、既に市川市で実用化されているものが挙げられる。これは、遠隔地の自動交付機における印刷によって、証明書類の交付を受けられるサービスである[3]。このサービスでは、まず市民が交付を受けたい証明書について、あらかじめ電子申請を行う。次に、該当する証明書が受け取り可能になると、メールにて通知が届く。受け取り可能メールの到着後、市民は27ヶ所のいずれかの自動交付機において証明書を受け取ることができ、さらに証明書を自動交付機で受け取ると交付手数料が50円割引になるという、利用促進のためのインセンティブも組み込まれている。

しかし、このモデルをそのままコンビニエンスストアにおける電子交付に当てはめることは難しい。それは、コンビニエンスストアにおける電子交付では、専用紙の利用が難しいという理由による。自動交付機では専用紙を用いることができるため、交付した証明書類の確からしさは、専用紙を用いているという事実による。しかしコンビニエンスストアにおける電子交付では、専用紙の管理が煩雑である等の理由により、専用紙を用いることが難しい。よって、汎用紙による電子交付が求められる。

汎用紙による電子交付では、印刷後の証明書類の確からしさを保証する手段として、コピー検知技術と改ざん検知技術が求められる。コピー検知技術は印刷物が別の印刷物

のコピーであるかどうかを検知するための技術であり、例えばコピーすると「複写禁」などの文字が浮かび上がるけん制模様や、専用装置によって観察可能かつ複写不可能な画像を印刷可能な特殊網点生成技術[4]が挙げられる。しかし、改ざん検知技術については、未だ十分には研究されたとは言えない。

本稿では、改ざん検知技術に適用可能な、印刷文書からの特徴抽出方式の考察について報告する。

2. 特徴抽出方式と関連研究

印刷文書の改ざん検知は、図1のように行われる。ただし、図1で言う「地紋」とは著者らが開発した地紋透かし技術[5]のことであり、

- ・ 特徴量を紙面に埋め込める
- ・ 視覚的にあまり目立たない
- ・ 地紋が本来印刷されているべき場所と、スキャンされた位置との差を認識可能なため、スキャン画像に含まれる紙の皺や折り目の影響を軽減させることが可能

といった特長を持っている。

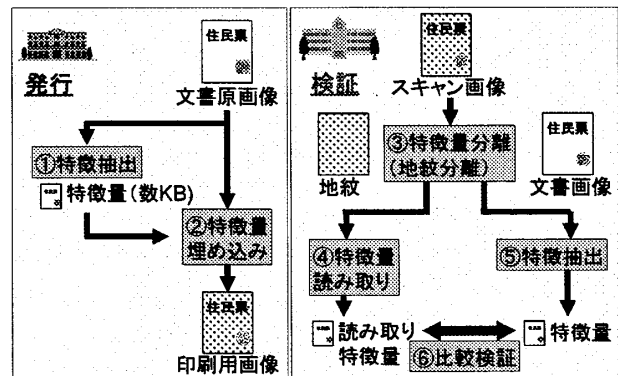


図1 印刷文書の改ざん検知

図1において、文書原画像から特徴を抽出しているが、その方式は以下の要件を満たす必要がある/満たしていると望ましい。

- ・ 非衝突性
- ・ 広範囲分布性
- ・ D/A, A/D変換における不変性

特徴抽出方式としては、文書中のブロックごとに、そのブロック内に含まれる黒画素の数、すなわち面積を特徴量とする方式が考えられる。しかしこの方式では、印刷のじみやスキャン時の2値化パラメータの影響によって、ス

(株) 日立製作所 システム開発研究所

キャン画像から算出した特徴量が、文書原画像から算出した特徴量と異なってしまうという課題が存在する。

著者らは上記のような課題を解決可能な、新たな特徴量抽出方式を検討している。

3. 重心利用型特徴量

著者らは、重心利用型特徴量抽出方式を検討している。これは、ある領域内に含まれる画素の重心を特徴量とするものであり、例えば図2左にあるように領域として「4」という文字一文字の領域を考えた場合、その重心である+記号で図示した位置の座標が特徴量となる。

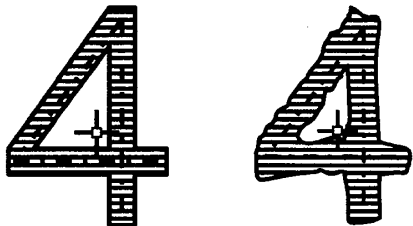


図2 重心利用型特徴量の模式図

この重心利用型特徴量の特長は、図2右に模式的に図示するようなにじみや、あるいはスキャン時の2値化の影響による文字の痩せ/太りが生じても、その影響を受けにくい点にある。これは、改ざん検出の主要な対象である文字が、ほぼ同程度の太さの線から成っているという性質を利用している。また、さらに太さの影響を排除するために、文字を細線化してから重心を算出することも検討している。

4. 評価

著者らは重心利用型特徴量の基礎評価として、重心利用型特徴量が、非衝突性や、広範囲分布性を持っていることについて評価した。

評価は、アルファベットおよび数字の重心位置を計測し、個々の重心位置同士の距離を測定するという実験によって行った。実験は一般的な英文セリフ書体にてアルファベット(大文字・小文字)および数字62文字を300DPI, 10ptにて描画し、その重心位置を文字ごとに計測した。また、比較のためにこれらの黒画素の量(面積)を計測した。

実験結果を図3に示す。実験結果は、横軸に特徴量同士の距離を、最大距離で正規化して表示している。また、縦軸はその距離関係の頻度である。なお、距離関係は62文字間の距離であるため、 $62 \times 61 / 2 = 1891$ 通りある。

全体として、重心による距離は、面積による距離よりも長いという傾向が見られる。これは、他の文字とより判別がしやすくなるため、好ましい。すなわち、他の文字との判別という観点からは、面積と比較すると重心位置の方がより好ましい特徴量と言える。

また本実験では、重心位置に関しては、距離=0、すなわち他の文字と同じ重心位置になる文字はなく、非衝突性を満たしていた。一方、面積に関しては全体の8%にあたる6文字が距離=0、すなわち他の文字の面積と同一の値とな

り、非衝突性が満たされていなかった。特徴量が衝突すると、文字を距離=0の他の文字と改ざんしてもその事実を確認できないため、特徴量として好ましくない。すなわち、非衝突性に関しても、面積と比較すると重心位置の方がより好ましい特徴量と言える。

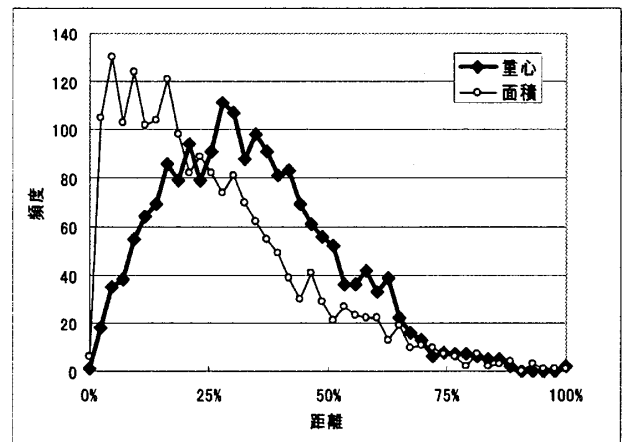


図3 重心位置の散らばり測定結果

5. まとめと今後の課題

本稿では、著者らが検討している重心利用型特徴量について、その概要を述べ、さらに基礎評価結果を示した。

重心利用型特徴量は、あるブロック内に含まれる黒画素の重心を、そのブロックの特徴量とするものである。基礎評価では、アルファベットや数字といった文字の重心間距離を計測し、面積と比較して、より衝突しにくいことや、広範囲に分布するといった、好ましい結果を示すことを確認した。

今後は、D/A, A/D 不変性の確認や、他のフォントを用いた実験を通じて、重心利用型特徴量抽出方式を確立し、文書の改ざん検知技術を確立していく予定にしている。

6. 参考文献

- [1] “重点計画-2006”, IT戦略本部, 2006.
- [2] “2007年度報告書”, IT新改革戦略評価専門調査会, 2008.
- [3] “電子交付サービスのご案内”, 市川市, 2008.
- [4] 長島寿人, “偽造防止を施した特殊網点生成技術”, 04-4印刷・情報記録・表示研究会, pp.5-6, 高分子学会, 2004.
- [5] 高橋由泰, 山田隆亮, 海老澤竜, 藤井康広, 手塚悟, “汎用プリンタ対応ドットパターン技術の研究開発”, FIT 2007, pp.325-326, 2007.
- [6] Yoshiyasu Takahashi, Takaaki Yamada, Ryu Ebisawa, Yasuhiro Fujii, Satoru Tezuka, “Information Embedding Method for Home Printing of Certifications”, In Proc. of International Conference on Advanced Communication Technology : ICACT 2008, pp.2116-2120, 2008.