

印字画線の劣化程度を指標とした電子写真方式プリンタの固体識別

Identification for laser printers using degraion of contours on printed characters

古川 猛[†]
Takeshi Furukawa

1. はじめに

犯罪に使用された脅迫文から印刷に使用した電子写真方式のプリンタを特定する新しい方法を提案する。法科学分野では、印刷文字を顕微鏡により拡大して観察することにより、微細な差異を検出してプリンタのメーカーや機種を識別が行われている。しかし、検査者の目視による観察では、差異の定量化が難しい。

文字認識の分野では、微細な差異を検出するためにウェーブレット分解を利用した方法が提案されている。Ding らは、ウェーブレットを漢字のフォントの識別に応用した[3]。信号処理の分野では、Mallat がウェーブレットにより検出したゼロ交差点が、数段階のスケールにおいて信号の形の変化や変化点を示すこと提言している。さらに、ゼロ交差点がパターン認識の問題点に適応できる表現と定義している[2]。我々も既にウェーブレットによるゼロ交差点を利用して電子写真方式のプリンタにより印刷した文字の微細な差異を検出した[1]。

前報では、わずか 3 機種であったプリンタの機種を本報告では、10 機種に増やしたため、機種識別が一層困難となった。この問題に対処するため、我々は前報と比較して二つの点を改良した。まず、一つ目は、ウェーブレット分解のスケールを増やした点である。次に、二つ目は、前報使用したゼロ交差点を用いなかった点である。前報では、一文字につき 5 段階のウェーブレット分解、及びゼロ交差点にて得られた 5 個の特徴ベクトルのみで、平均 2 乗誤差によりプリンタの機種識別が可能だったものが、今回は機種を増やしたため、ウェーブレット分解によって得られたプロファイル全体を使用し識別する工夫が必要となった。この報告では、提案手法が、電子写真方式のプリンタのメーカーや機種を識別に有効なことを示す。

2. 提案手法

この章では、まず、多段階のスケールのウェーブレット分解による特徴抽出について示す。次に、幾つかの部分空間法に基づいた認識方法の詳細を記載する。図 1 に提案手法のおおまかな処理の流れを示す。

2.1 電子写真方式のプリンタにより印刷された文字の輪郭線へのウェーブレット分解の応用

この章では、電子写真方式のプリンタにより印刷された文字から特徴抽出の過程において、ウェーブレット分解を適用した点を述べる。ウェーブレット分解は、しばしば信号や画像の処理、時系列データの処理に使用される。特に、実際の信号は、粗い成分から細かい成分まで幾つかのスケールの情報含んでいる。ウェーブレット分解は、それらの信号からさまざまな異なったスケールに

おける情報を抽出できる。この点に注目して、ウェーブレット分解をプリンタ文字の識別に適用するために、プリンタ文字の輪郭線を信号として扱う。

電子写真方式のプリンタにより印刷された文字を観察すると、図 1 に示すように文字の輪郭線の品質に差異があること解る。この差異は、プリンタメーカーが、プリンタの開発や製造に際しての機能の実装にあたり、どの点に重点を置くか、例えば、線の端正さ、印刷速度、トナーの消費量などにより生じる。さらに、印刷用紙の品質、すなわち紙繊維の影響、コートの有無などの紙の表面処理の影響、加えて、印刷ドラムや現像処理に係る消耗品の劣化の程度の影響による。提案手法の詳細を以下に示す。Ding ら]の先行研究[3]を参照して、直交ウェーブレットを今回の実験に使用した。ウェーブレット分解の段階は、7 段階とした。よって、一つのデータセットは、7 個の波形のベクトルから構成される。最も長い第一段階の波形の長さに基づいて各波形の長さは 2966 とした。前報では、わずか 3 機種のプリンタの分類であったため、5 段階のウェーブレット分解後の五つの波形からゼロ交差点を使用して特徴量とした。さらに、識別では平均 2 乗誤差を使用した結果、正解率は 0.7 であった。

しかし、本報では、プリンタを 10 機種に増やしたため、前報の方法では、機種識別が不能となった。よって、今回は、ゼロ交差点の数では識別に必要な特徴量が不足するため、ゼロ交差点は使用しなかった。ゼロ交差点の代わりに、全波形を使用したため、特徴ベクトルの数は、1 から 2966 に極端に増加した。さらに、波形の識別には平均 2 乗誤差は適していないため、同法も使用できなくなった。よって、プリント文字の輪郭線からウェーブレット分解により得た全ての分解波形を使用して、印刷に使用したプリンタを識別するためには、もっと洗練された方法が必要となった。次の章では、この課題を克服するために採った方法を紹介する。

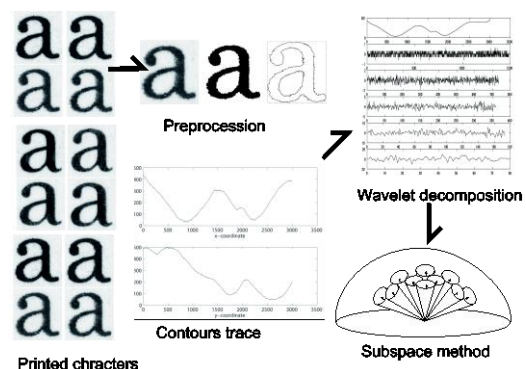


図1 提案手法の流れ

[†] 茨城県警察本部科学捜査研究所

Forensic Sci. Lab. Ibaraki Pref. Police H.Q.

2.2 部分空間法に基づいたプリンタ識別

先の章で述べたように、プリンタ文字からプリンタの機種を識別するために適した方法が必要となった。この課題に対処するため、部分空間法を使用した。部分空間法は、3次元の物体識別[4]や手指[5]や顔認識に広く使用されている。まず、学習過程では、サンプルデータの自己相関係数が計算される。識別する対象と学習データの類似性は、対象から得たベクトルと学習データの固有ベクトル間の内積を計算することにより評価される。類似性は、入力ベクトルと部分空間の成す角度で決定される。相互部分空間法では、学習データの部分空間と同様に、複数の入力ベクトルから作成した部分空間を作成する。直交相互部分空間法は、クラスの識別能力を向上させるために辞書の部分空間間の関係を直交化する過程を含んでいる。

入力ベクトルのクラスの数が増加するにつれ、各クラス間の類似性も増していき、そのため、識別の正解率が減少する。この弱点を克服するため、相互部分空間法では、辞書空間におけるクラス間の関係を直交化した。直交相互部分空間法は、高い識別力を示すものの、クラス間の関係が非線形である際は、認識率が低下する。このような非線形なデータに対応するため、近年、福井らは、カーネル直交相互部分空間法を提案した[4]。本報告においても同法を使用した。

表 1 実験に使用した電子写真方式プリンタ

メーカー	機種	方式
Ricoh	SP8100	Laser
Ricoh	SPC710e	LED
Ricoh	SPC301e	Laser
Fuji Xerox	DPC3250C	Laser
Fuji Xerox	DCC5580	Laser
Canon	iR4045V	Laser
Ricoh	DPP350d	Laser
Epson	LPS6500	Laser
NEC	MW8400N	Laser
NEC	MW8250N	Laser

3. 実験と考察

3.1 評価システム

まず、実験材料について述べる。実験に使用した 10 機種の電子写真方式のプリンタを表 1 に示す。各プリンタを使用して英語のアルファベット 26 文字種を一度に 10 枚の用紙に連続して印刷した。文書作成に使用したフォントは、センチュリーで、実験に使用した文字種は、英語の小文字の「a」である。印刷文字の撮影にはフラットベッド型イメージスキャナ(iQ smart 3, Kodak (Creo))を使用した。撮影した解像度は 5,400dpi、諧調は 256 段階とした。

3.2 前処理

図 1 に前処理の例を示す。高解像度スキャナにより撮影された原画像は、Otsu 法により 2 値化された。その後、2 値化された文字画像から輪郭線を抽出した。輪郭線画像

は、図 1 に示すように x 軸と y 軸毎に座標値として表現された。その後、両軸の座標値を結合した。

3.3 メーカー機種識別実験

表 1 に、実験に使用したプリンタを示す。各プリンタは、英語の小文字「a」を 10 枚の用紙に連続して印刷した。印刷は、4 日に分けて行った。よって、10 機種×4 日×10 回の計 400 個の文字サンプルが収集できた。同じ機種のプリンタにより同じ日に印刷された文字を一つセットとした。その後、同じ機種により印刷され、しかも印刷日が近いセットを組み合わせた。よって、印刷日が離れた 2 組のデータセットができた。一組のデータセットは、10 機種×10 回×2 日の 200 文字で構成される。最初の一つのデータセットを参照データとし、もう一方のデータセットをテストデータとした。参照データによる辞書の学習過程では、20 個文字から選択する先頭の文字を 1 個ずつ順にスライドさせ、10 個を選択して辞書を作成した。

表 2 に機種の識別実験の結果を示す。実験における部分空間の次元は、参照データの作成過程では 5 段階、テスト過程では、3 次元とした。誤り率は、相互部分空間法、カーネル相互部分空間法、カーネル直交相互部分空間法の順で減少した。カーネル直交相互部分空間法を使用した場合、最も良い識別結果を得た。

4. 結論

本報は、電子写真方式のプリンタにより印刷された文字から印刷に使用したプリンタを識別する新しい方法を提案した。本提案手法は、法科学における電子写真方式のプリンタの識別に有効だけでなく、印刷した文字の品質を評価することによりプリンタの維持管理にも有効である。今後、新品の電子写真方式のプリンタを購入し、印刷枚数を定期的に記録することにより、さらに正確な実験を試みる。

表 2 実験結果

	SM	MSM	KMSM	KOMSM
ER	0.5536	0.3636	0.3364	0.2364
EER	0.3127	0.2465	0.2727	0.1727

謝辞

本研究は科研費奨励研究(H270919)の助成によるものである。部分空間法の適用について、筑波大学システム情報工学研究科福井和広教授に貴重な御助言をいただいた。

参考文献

- [1] T. Furukawa, A new method for discriminating printers based on contours qualities of printed characters using wavelet decomposition The proceedings of ICDAR2013, pp.1147–1151, (2013).
- [2] X. Ding, L. Chen and T. Wu. Character independent font recognition on a single Chinese character, IEEE Trans. Pattern analysis and Machine Intelligence vol. 29, No.2, pp.195–204, (2007).
- [3] S. Mallat, Zero-crossings of a wavelet transform, IEEE Trans. Inf. Theory, vol. 37, no. 4, pp. 1019–1033 (1991).
- [4] K. Fukui and O. Yamaguchi, The kernel orthogonal mutual subspace method and application to 3D object recognition, Proceedings of Asian conference on computer vision, pp.467–476, (2007).
- [5] Y. Ohkawa and K. Fukui, Hand shape recognition based on kernel orthogonal mutual subspace method, Proceedings of conference on Machine Vision Applications, pp122–125, (2009).