

機械学習による印字からのプリンタ印刷方式の識別

Recognition of Printing Methods from Printed Characters Using Machine Learning

古川 猛

Takeshi Furukawa

1. はじめに

近年、画像処理技術、微細加工技術の発展により、家庭用プリンタの印刷品質の向上がめざましい。特に、インクジェット方式複合機を始めとするプリンタ類は、価格も安く、装置も小型のため、家庭への普及は著しい。それに伴い、同プリンタ類による偽造事案も多発している。そういった状況下の犯罪捜査において、偽造に使用した機器を特定する事は、犯罪の裏付けとして大切である。非破壊検査として、我々は色材の付着状態を顕微鏡による拡大観察にて識別する方法を採っている。法科学における文書鑑定人によれば、拡大観察によりインクジェット、あるいはレーザー等の印刷方式を識別することが可能である。また、機種毎の微細な違いを観察することもできる。しかし、専門家によれば真偽の見分けは簡単であるものの、警察官や一般人にとっては荷が重い。そこで、簡単に区別するための支援システムを開発することを目標とする。今回は、識別に用いる手法の性能を評価するための基礎実験を行った。

前報では、トナーを色材として使用する電子写真方式のレーザープリンタ類を対象とした方法を提案した[1]。現在までに、レーザープリンタのメーカー、機種、印刷枚数により文字の外観に差が生じる事を、新規に 8 機種種のプリンタを購入した上で、未使用の状態から印刷を始め、1 から 1000 ページの印刷サンプルを作成し、このサンプルを使用する際の識別実験にて劣化による差異も確認している[2]。今回、さらに、インクジェット方式のプリンタ類による印字サンプルを前報の実験に追加した上で、機械学習により印字からプリンタの印刷方式が比較可能か検討した。

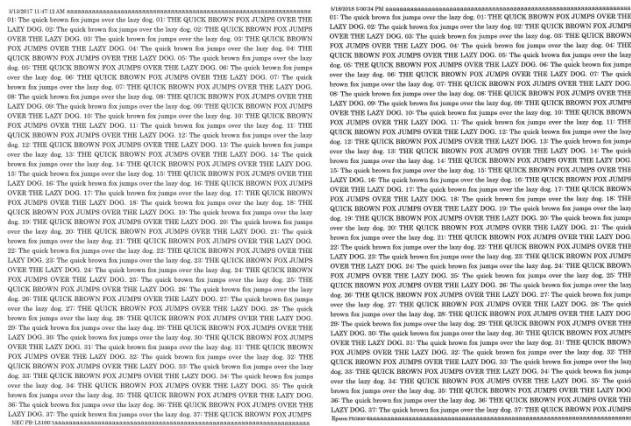


図 1 実験に使用したサンプル例

左:レーザー(日本電気)、右インクジェット(エプソン)

茨城県警察本部科学捜査研究所, Forensic Science Lab.
Ibaraki Pref. Police H.Q.レーザー
プリンタインクジェット
プリンタ

図 2 実験にて使用した印字の例

2. 方法

提案する方法において、特徴抽出にはウェーブレット分解、識別には部分空間法を使用した。

2.1 手法

印字画像からの特徴抽出にあたっては、ウェーブレット分解を使用した。同分解により多段階のスケールごとのプロファイルを得て特徴量とした。識別には、機械学習の手法である部分空間法を使用した。以下に詳細を示す。

2.1.1 ウェーブレット分解

前報[1, 2]と同様に文字の輪郭を一次元の信号として扱い、特徴抽出には、ウェーブレット分解を使用した。ウェーブレットは、信号や画像からの特徴抽出に使用される。特に、マルチスケールウェーブレットにより、さまざまスケールごとの特徴、例えば、高周波のノイズの検出や瞬間的な振動の抽出が可能である[3]。文字認識関係の分野では、Dingら、ウェーブレットを漢字のフォントの識別に応用した[4]。インクジェットやレーザープリンタにより印刷された文字についても、さまざま要因が文字の印刷品質に関係していると考えられウェーブレットによりそれらを検出する。

2.1.2 部分空間法

部分空間法は、顔画像や物体認識に広く使われている[5]。その長所は、同じ機械学習の手法の一つであるサポートベクターマシン (SVM) と比較して、多クラスの識別が容易である点あげられる。なお、SVM と同様に部分空間法においても、カーネル直交相互部分空間法 (KOMSM) などの非線形データに対応するカーネル法を使用できる。さらに、ディープラーニング等のニューラルネットワークと比較して、数学的定義が明確である。以上の点から、今回の二方式の複数のプリンタの識別に各種の部分空間法を使用した。

実験

提案した方法を評価するため以下の実験を行った。以下に詳細を示す。

2.1.3 機材

まず、実験に使用した機材を示す。表 1, 2 に実験に使用したプリンタを示す。各プリンタで白色紙 (極東、Premium white PPCKA4) に印刷した。図 1 に印刷した用紙の一例を示す。印刷した文は、「The quick brown fox jumps over the lazy dog. THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.」であり、今回、分析に使用した文字種は、「a」字を使用した。1 ページには、37 回繰り返して上の文が印刷されており、ヘッダーとフッターにも「a」字を印刷した。サンプルの印刷に際しては、マイクロソフト社ワード 2010 を使用し、フォントはセンチュリー、大きさは 12 ポイントとした。各ドライバーの印刷品質の設定は、標準とした。

文字の撮影には、測定顕微鏡(ミットヨ MF-UD4020D)に、画素ずらし方式の CCD カメラ (JENOPTIK ProgRes C14 plus) を装着して行った。照明は、暗視野照明法とした。処理した文字画像の大きさは、2401×2401 画素、階調は 256 段階とした。図 2 に実験にて印刷した文字の例を示す。

表 1 実験に使用したレーザープリンタ

メーカー	機種
ELECOM	EPR-LS01W
CANON	LBP6030
NEC	PR-L5100
BROTHER	HL-L2300
KONICA MINOLTA	PagePro 1350W
FUJI XEROX	DocuPrint P250 dw PCI 6
OKI	B801
RICOH	IPSio SP3400L

表 2 実験に使用したインクジェットプリンタ

メーカー	機種
CANON	MP980
EPSON	PX-G930
EPSON	PX-7V

2.1.4 前処理

文字画像を Otsu 法により 2 値化した[6]。その後、輪郭線を抽出した後、x 軸と y 軸別に輪郭線の方向を追跡し、波形とした。

2.1.5 特徴抽出

画像から抽出した波形を 8 レベルでウェーブレット分解を行った。使用したウェーブレットは、Ding らの先行研究[4]を参考として直交ウェーブレットを使用した。

2.1.6 識別

識別には、部分空間法 (SM) を使用した。使用した部

表 3 実験結果：レーザー8機種、インクジェット3機種

	SM	MSM	KMSM	KCMSM	KOMSM
ER	0.0913	0.0141	0.0090	0.0077	0.0064
EER	0.0816	0.0166	0.0346	0.0228	0.0038

分空間法は、相互部分空間法 (MSM)、カーネル相互部分空間法 (KMSM)、カーネル制約差分相互部分空間法 (KCMSM)、カーネル直交相互部分空間法 (KOMSM) とした。まず、紙面上に印刷された 100 文字を撮影した。次に、同 100 文字から無作為に選んだ 20 文字を学習サンプルとし、残り 80 文字をテストサンプルとして識別実験を行った。テスト試行では、80 文字から 10 字を 1 字ずつ、ずらして取り出した 11 機種 71 セットを比較した。

3. 結果

識別結果の一例を表 3 に示す。実験における部分空間の次元は、学習では 10 次元、テストでは 4 次元とした。使用したカーネルは、ガウスクERNELを使用し、ガウシアンカーネルパラメータは 1 とした。図 3 左列に偽陰性率 (FRR)、偽陽性率 (FAR) 曲線を示す。試行を繰り返す中で、誤認識率は低下した。図 3 右列に機種毎の結果を示す。レーザープリンタ間では、7 件の誤識別があったものの、レーザーとインクジェット間では誤識別はなかった。よって、本手法の有効性が確認できた。

今後は、さらに、機種、印刷枚数を増やし、特に、インクジェットプリンタの劣化の影響を調べ、支援システムへの実装を計る。

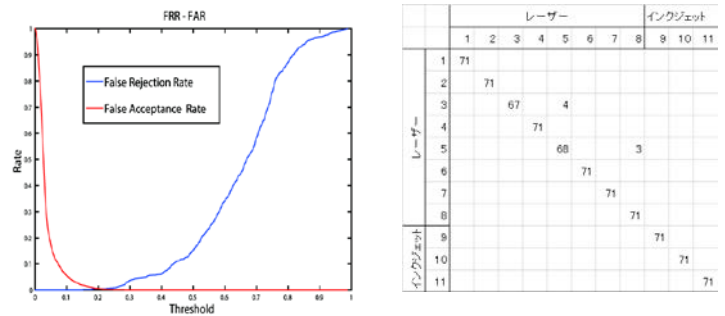


図 3 実験結果 左:FRR-FAR 曲線、右:混同行列

謝辞

部分空間法について筑波大学システム情報工学研究科福井和宏教授に指導をいただいた。本研究は JSPS 科研費 JP18H00548 の助成を受けたものです。

参考文献

- [1] T. Furukawa, "A new method for discriminating printers based on contours qualities of printed characters using wavelet decomposition", Proceedings of ICDAR2013, (2013).
- [2] T. Furukawa, "Subspace based method with multi scale wavelet for recognition of printer property", Proceedings of ICDAR2015, (2015).
- [3] チャールズ K.チュウイ, "ウェーブレット応用", 東京電機大学出版局, (1997).
- [4] 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 (2007).
- [5] K. Fukui, O. Yamaguchi, "The kernel orthogonal mutual subspace method and application to 3D object recognition", Proceedings of Asian conference on computer vision, (2007).
- [6] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Systems, Man and Cybernetics, Vol.9, No. 1 (1979).