

機械学習によるレーザープリンタにより印刷された文字のメーカー、機種 of 識別 Recognition of models and manufactures of laser printed characters using machine learning

古川 猛^{†‡}
Takeshi Furukawa

1. はじめに

ストーリーによる中傷ビラの配付などの犯罪捜査において、文書を印刷した機器を特定する事は、犯罪の裏付けとして重要である。特に、電子写真方式のプリンタ、いわゆるモノクロレーザープリンタは、印刷速度が速いことから大量の中傷ビラの印刷に使用されることが懸念される。法科学の分野では、印刷された文字を形成する黒色トナーの化学的分析によりメーカー、機種を特定している。しかし、これらの方法は、文字を形成するトナーを削り取る破壊検査が必要である。一方、非破壊検査としては、確立した方法がないため、我々はトナーの付着状態を拡大観察して識別する方法を提案している[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確認している[2]。しかし、実験において、使用した文字のサンプルが、我々の研究室で日常使用しているレーザープリンタにより印刷したものであるため、印刷枚数による文字の劣化を定量的に統制した実験とは言いがたい。よって、今回、新規に 8 機種のプリンタを購入した上で、未使用の状態から印刷を開始し、1 から 1000 ページの印刷サンプルを作成した。このサンプルを使用して、識別実験を行ったので報告する。



図 1 レーザープリンタにより印刷したサンプル例 (ELEC1000 枚目)

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、識別には部分空間法を使用した。

2.1 手法

文字の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、機械学習の一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 らが、ウェーブレットを漢字のフォントの識別に応用した[4]。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[5,6]。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。以上の点から、今回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。



図 2 実験にて印刷した文字の例
左上 CANON1 枚目 左下 CANON1000 枚目
右上 BROTHER1 枚目 右下 BROTHER1000 枚目

[†] 茨城県警察本部科学捜査研究所, Forensic Science Lab, Ibaraki Pref. Police H.Q.

[‡] 筑波大学システム情報工学研究科, Graduate School of Sys. & Inf. Engineering, Uni. of Tsukuba

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 実験結果：1 枚目と 2 枚目

	SM	MSM	KMSM	KOMSM
ER	0.4878	0.4726	0.5457	0.3841
EER	0.2104	0.3476	0.3081	0.1618

表 3 実験結果：999 枚目と 1000 枚目

	SM	MSM	KMSM	KOMSM
ER	0.6585	0.6250	0.5945	0.4848
EER	0.2774	0.3754	0.3354	0.2104

3. 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。

3.1.1 機材

まず、実験に使用した機材を示す。表 1 に実験に使用したレーザープリンタを示す。各プリンタで白色紙 (極東、Premium white PPCKA4)1000 枚に印刷した。印刷に際しては、予め各プリンタに装着された純正なトナーカートリッジを使用して、トナー切れで印刷ができなくなるまで印刷した。1000 枚までにトナーが切れたため、各プリンタの純正トナーカートリッジに交換し印刷を続けた。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。各ヘッダーには印刷日時、各フッターには、メーカー、機種、印刷カウント数が印刷され、8 機種×1000 枚=8000 枚のサンプルを 1 枚ずつ区別できる。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(ミツトヨ MF-UD3017B)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。

3.1.2 前処理

文字画像を Otsu 法により 2 値化した[7]。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。最後に両軸から抽出した波形を結合した。

3.1.3 特徴抽出

画像から抽出した波形を 10 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。

3.1.4 識別

識別には、部分空間法 (SM) を使用した。使用した部分空間法は、相互部分空間法 (MSM)、カーネル相互部分空間法 (KMSM)、カーネル直交相互部分空間法 (KOMSM) とした。まず、1 枚目に印刷された 50 文字を学習サンプルとした。次に 2 枚目に印刷された 50 文字をテストサンプルとして識別実験を行った。同様に 999 枚目に印刷された 50 文字を学習サンプルとし、1000 枚目に印刷された 50 文字をテストサンプルとして実験を行った。

4. 結果

識別の結果を表 2、3 に示す。実験における部分空間の次元は、学習では 3 次元、テストでは 3 次元とした。誤り率は、カーネル直交相互部分空間法 (KOMSM) を使用した際が、最も低かった。同法は、非線形なデータの識別にも有効であることから、レーザープリンタにより印刷された文字には、非線形な要因による形状の変化が生じていることが考えられる。

ヒトの目視における観察では、印刷による文字画像の劣化は、1、2 枚目よりも 999、1000 枚目に顕著である。一方、今回の機械による識別では、999 枚目と 1000 枚目間の比較が、1 枚目と 2 枚目の比較よりも誤り率が高い結果となった。印刷枚数を増すにつれ文字画像が劣化し、その影響により識別が困難になるか、どうかは、今後、さらに、500 枚目などの中間のページの文字画像を解析して、原因を探る。

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。

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] Y. Ohkawa, K. Fukui, "Hand shape recognition based on kernel orthogonal mutual subspace method", Proceedings of conference on Machine Vision Application, (2009).
- [7] N. Otsu, "A threshold selection method from gray-level histograms", IEEE Trans. Systems, Man and Cybernetics, Vol.,9, No. 1 (1979).