

研究指導 中澤 真 教授

自由手書き文字の認識精度低下の要因分析

池谷 友宏

1. はじめに

情報技術やインターネットの発達により様々なものが電子的に管理され、共有することが容易になった。そのため、ビジネスや学校などで使われる文書も電子化する動きが活発になっている。しかし、元々紙媒体の文書を電子化のために手入力で作成しなおすことは、莫大な労力を必要とし効率的ではない。そこで活躍しているのが画像から文字を読み取り、テキストデータ化するOCR¹と呼ばれる技術である[1]。この技術により文書上に書かれた情報を電子的に扱えるようになり、またテキストデータとしても保存されているため端末上での検索が可能になった。

しかし、文書の電子化に用いられる文字認識ソフトは活字を対象としたものが多く、手書き文字に対しては、認識精度が低くなってしまふ。手書き文字に対応した文字認識ソフトもあるが高精度の認識率を謳うものは帳票や記入枠などのフォーマットが限定された書類の手書き文字にしか対応していない。一方、ノートやホワイトボードなどのフリーフォーマットで書かれる自由手書き文字に対しては、専用のソフトはないため一般的な文字認識ソフトを使用するしかなく、低い認識精度しか達成できない。このように、現状では活字の文字認識と比較して自由手書き文字の認識はかなりの課題を抱えている。この原因として自由手書き文字にみられる個人の癖や大きさのばらつきが挙げられる。よってこの特徴を除去することで認識精度が向上する可能性がある。

そのため本研究では自由手書き文字に表れる特徴が認識精度低下に与える影響を明らかにし、文字画像からこれらの特徴を取り除く前処理を施すことによる認識精度向上を目指す。

2. 文字認識技術の概要

一般に文字認識はオンライン文字認識とOCRの2つに大きく分けられている。オンライン文字認識はスマホなどの文字入力の場面で使われており、文字の書き順情報も利用可能である。一方、ノートや紙媒体の書類上の文字をカメラやスキャナでデジタル画像として取り込み、これをテキストデータ化するのに用いられるのがOCRである[2]。OCRの場合、すでに書き終わった文字の認識となるため書き順情報は得られない。よって文字の形のみで認識するので、その

精度向上のために活字認識と手書き文字認識それぞれで研究が進められている。

活字を対象とした文字の認識率は98~99%の精度を誇り、残る課題は100%に近づけるだけだといわれている[3]。一方、「DX Suite」[4]というソフトでは手書き文字の認識精度も活字と同様に高い。しかし、これらの手書き文字認識は帳票や記入枠に書かれた文字の認識であり、日常生活で使われる自由手書き文字の認識精度は低い[5]。

2.1 OCRに必要な文字の画像化

OCRには文字画像が必要であり、今まで文書はスキャナで画像化することが多かった。現在ではデジカメやスマートフォンの普及により手軽に画像化でき、以前よりも効率的なOCRが可能となった。

ただし、画像化に使用する機材によっては認識精度に影響を及ぼす可能性がある。実際に黄瀬ら[6]の研究ではデジカメで撮影した文字画像を使用した文字認識の精度を比較している。結果として撮影する際の手振れや明度、解像度による劣化の影響により、認識精度が低下することを明らかにしている。しかし、2005年当時のデジカメを使用した研究であるため現在主流のスマートフォンのカメラで撮影された画像を用いた場合、認識精度にどのような差異が生じるかは明らかになっていない。

2.2 自由手書き文字認識の現状と課題

現在の文字認識は活字だけでなく、「DX Suite」のようなソフトや佐々木ら[7]の研究のように手書き文字も認識できるようになっている。佐々木らの研究では畳み込みニューラルネットワーク(CNN)を用いることで高い認識精度を出しているが、使用された文字画像は1文字ずつに切り出されており、自由手書き文字の画像を用いた検証はしていない。ビジネス分野では帳票のような様式が決まった書類が多いのは確かだが、会議にホワイトボードを使用したり、ノートや付箋などのフリーフォーマットの書類も多く存在し、自由手書き文字に対する文字認識を必要とする場面も少なくない。プライベートで利用する場面では、より自由手書き文字認識が必要となるだろう。

しかし、自由手書き文字は書き手の癖が顕著に表れ、形が個人によって大きく異なる。そのうえ枠が設けられていない場に制限なく書かれるため、自由手書き文字の認識は一般の文字認識より精度が低くな

¹ <https://qiita.com/icoxfog417/items/5fd55fad152231d706c2>

る。ゆえに、認識する画像を定型フォーマットに近い、文字が均一に書かれた状態へと前処理してから文字認識することが重要になる。

以上のことを踏まえて本研究では自由手書き文字に表れる、活字にはない個々の文字の大きさや傾きなどのばらつきが認識精度低下に与える影響を分析する。認識精度に影響を及ぼす自由手書き文字固有の特徴を明らかにすることで、これらの特徴を打ち消す前処理による認識精度向上の可能性を探る。

3. 精度低下の要因分析のための実験

3.1 実験概要

自由手書き文字で書かれた文書を全文検索可能なテキストデータへと変換することは現状の技術レベルでは難しい。認識精度向上のための第一歩としては、先に述べたように認識前に文字画像へ前処理を施すことが重要である。そこで、どのような前処理が有用か明らかにするために、自由手書き文字を画像化した際に表れる以下の特徴が、認識精度にどのような影響を及ぼすか実験する。

- 撮影機材による差
- 文字色による差
- 背景色による差
- 文字の大きさによる差
- 文書全体の傾きによる差
- 文字の傾きによる差
- 行間の広さによる差
- 文字間隔の広さによる差

実験を行う際に必要となる文字認識処理はGoogleから提供されている「Cloud Vision API」²を利用した。それを呼び出すプログラムをPython[8][9]で作成して文字認識を行った。また文字画像の撮影にはGoogle社のスマートフォン「Pixel3」とEPSONのスキヤナ「GT-X830」を使用した。

画像名	総文字数	画像名	総文字数
活字画像1	126	活字画像9	535
活字画像2	887	活字画像10	488
活字画像3	171	活字画像11	791
活字画像4	341	活字画像12	242
活字画像5	612	活字画像13	347
活字画像6	1121	活字画像14	128
活字画像7	1208	活字画像15	180
活字画像8	756		

表 1 実験用文書数と総文字数の一覧

実験には手書き文字ではなく活字を使用する。これは活字にすることで手書き文字一つひとつの癖の差異による認識精度のばらつきを要因から排除し、先に挙げた要因候補のみに注目して認識精度低下への影響を明らかにするためである。なお、活字画

像はWordにて作成した15種類の文書を印刷し、これを撮影した画像データを使用する。また活字文書は文の構成が偏ることを防ぐため、総文字数をランダムにした。詳細は表 1のとおりである。

3.2 撮影機材による影響

一般的に、文字の画像化にはスキヤナを用いることが多いが、携帯性はないため手軽に画像化できない。またホワイトボードに書かれた文字はスキヤナで読み取ることができないためカメラによって画像化するしかない。この画像化に使用されるデジタルカメラやスマートフォンのカメラは、手振れ防止機能の有無や解像度特性など撮影機材の差異があり、これが画質に影響を及ぼす可能性を考えなければならない。

そこで、異なる機材で撮影された文字画像の認識精度の差異について検証した。ここでは、手振れの心配がなく高画質なスキヤナと、広く利用されているが手振れの可能性があるスマートフォンカメラの二種類の機材を使用して実験をした。なお、文字認識精度を表す認識率は以下の式で算出している。

$$\text{認識率} = \frac{\text{正しく認識された文字数}}{\text{総文字数}}$$

ここで、認識の正誤は元データと認識後のテキストデータの差分処理³によって判断し、総文字数から差分処理で計算された削除文字数を引いた数を「正しく認識された文字数」としている。

この実験の結果、スキヤナで画像化した場合に15種類の活字画像の平均認識率は99.9%、スマホカメラで画像化した場合の平均認識率は99.8%とほぼ変わらないということがわかった。これは文字を画像化する撮影機材の差異が、認識精度に影響を及ぼさないことを示している。スマートフォンカメラの性能が上がっており、撮影画像の品質がスキヤナで読み込んだ場合とほぼ変わらず、認識精度に影響を与えるほど劣化しないためであると考えられる。

3.3 カラーによる影響

黄瀬らの研究では認識精度低下の要因として文字色について言及している。そこで、文字色と背景色それぞれについての認識率への影響を検証した。

3.3.1 文字色による差

実験に使用した画像は活字文書4つで、文字色が黒一色、赤一色、黒赤混合の3つの条件で認識率を比較した。結果は3つの条件すべての平均認識率が99.7%と同率であるため、文字色の差異は認識精度に影響しないことが明らかになった。

3.3.2 背景色による差

ノートやメモに書かれた文字はマーカーによって色付けされる場合がある。その際にマーカーの色が認識率に与える影響を明らかにするための実験を行った。実験には文字が黒で、1行間隔で黄色の

² <https://cloud.google.com/vision/?hl=ja>

³ <http://lab.hidetake.org/diff/>

マーカーが引かれている5つの活字文書を使用した。その結果、平均認識率は99.8%と高精度であり、マーカーによる色付けは認識に影響を与えないことが示された。以上の実験から色情報を削除する前処理は認識精度向上に寄与しないということが明らかになった。

3.4 文字の大きさによる差

手書き文字は活字のように文字の大きさが一定になることはなく、画数が多い漢字などはひらがなやカタカナに比べて大きくなりやすい。そのため、一部の文字が大きくなっている画像を認識した際に認識率にどのような影響を与えるのか実験した。

実験には文字の大きさが均一ではなく、一部に大きい文字が混在している図 1のような画像を5つ用いて認識率を比較した。なお、フォントサイズを変更する文字は単語単位で各行に1箇所とし、その位置はランダムに決めて偏りが生じないようにしている。また、フォントは游明朝を使い、ベースフォントサイズを10.5ptとし、これに対して14, 20, 24ptの3種類の変化をさせた場合の認識率を求めた。

この結果を図 2に示す。なおグラフは5つの文書、各フォントサイズの誤読率の平均を表したものである。また誤読率は以下の式で算出した。なお、誤読数は差分処理で算出された削除文字数である。

$$\text{誤読率} = \frac{\text{誤読数}}{\text{変更箇所の文字数}}$$

図 2からフォントサイズが増加するごとに平均誤読率も高くなることがわかる。そのため手書き文字認識を行う際は文字の大きさをそろえる前処理をすることで認識率が向上すると考えられる。

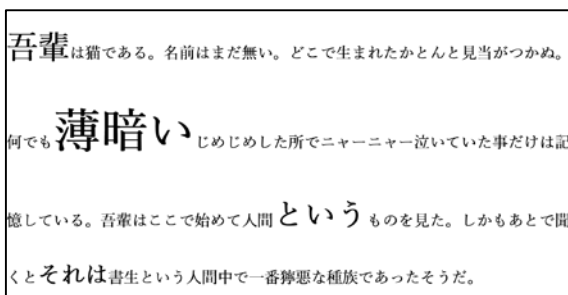


図 1 使用した文書の例

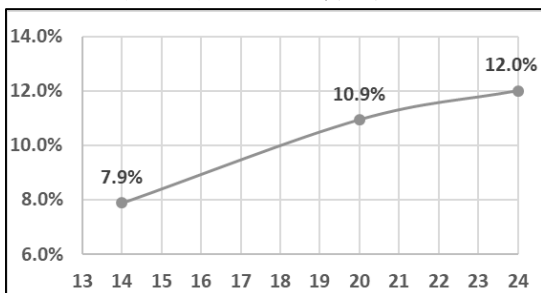


図 2 文字の大きさによる平均誤読率 (縦軸: 誤読率 横軸: フォントサイズ [pt])

3.5 角度による差

ホワイトボードなどに文字を書く場合、活字のように水平に書き続けることは難しく、一部の文字が水平方向に対して傾いてしまうことがよくある。また、スキャナに置く読み取り原稿がずれていたり、撮影に使用するカメラが水平を保っていない場合には文書の並びは傾いてしまう。そのため、それらが認識精度に与える影響を明らかにする実験をした。

3.5.1 文書全体の傾きによる差

まず文書自体が傾いている場合の認識率の変化を検証した。実験は5つの文書を作成し、全体の傾きを5, 10, 20, 30, 40度の5種類に変化させた文書の画像25枚を使用した。角度ごとの5枚の画像の平均認識率をグラフにしたものを図 3に示す。

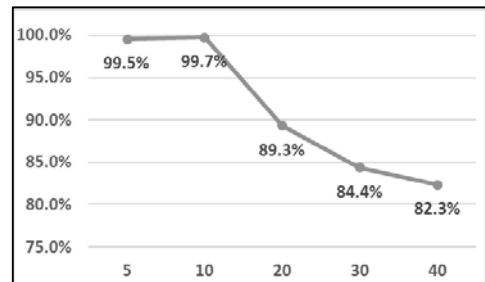


図 3 文書全体が傾いた場合の平均認識率の変化 (縦軸: 認識率 横軸: 角度 [度])

この結果より文書全体の傾きが増加するごとに平均認識率が低下していくことが見て取れる。そのため文書の角度を水平にするような前処理をすることで手書き文字の認識精度の向上が見込めることが明らかになった。

3.5.2 文字の傾きによる差

続いて文書中に傾いている文字が混在している場合の実験をする。文書を5つ作成し、文字を1, 2, 3, 4, 5度と5種類の変化をさせた文書の画像25枚を用いて実験をした。なお、傾きをつけた文字は3.4と同様に単語単位で各行に1箇所とし、位置はランダムに決めた。また、誤読率も前節と同様に算出した。実験結果は図 4のとおりである。

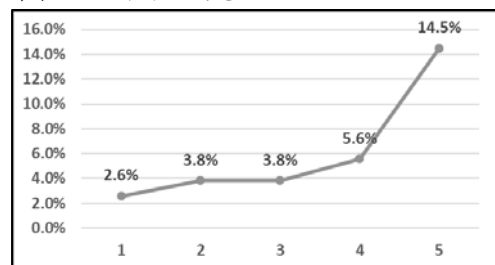


図 4 文字の一部が傾いた場合の平均誤読率の変化 (縦軸: 誤読率 横軸: 角度 [度])

上図は角度ごとの5枚の画像の平均誤読率を表している。結果として傾いている文字の角度が増加するごとに平均誤読率がわずかに上昇していることがわかる。このことから、文字が水平になる前処理を施すことで認識率が向上することが示された。

3.6 行間の広さによる差

文字認識を行う際に文と文の行間が狭いと文字が重なり、認識精度が低下してしまう。そのため文字認識ソフトの多くは行を切り離すような処理をしている。

しかし行間調整をするのに、行間を広げ過ぎた場合の認識精度への影響を考慮しないと適切な前処理ができない可能性がある。この影響を明らかにするためにWordにて5つの文書を作成し、各文書の行間を1.15, 1.5, 2, 2.5, 3ptと変化させた文書画像25枚を用いて実験をした。その結果が図 5である。

このグラフをみてわかるように行間を広げていった場合に平均認識率の低下はほとんどない。よって行間を広げる前処理をする際は、その広さの上限を考える必要はないといえる。

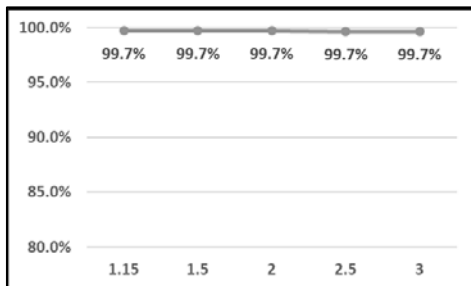


図 5 行間ごとの平均認識率の変化
(縦軸: 認識率 横軸: 行間の広さ[pt])

3.7 文字間隔の広さによる差

文字認識において重要となるのが文字と文字の間隔である。前述したとおり文字同士が接触していたり、重なっていたりすると認識精度は低下する。ただし行間の処理と同様に文字間隔を離す処理を考える際には、文字と文字の間隔を大きく広げ過ぎたことが、認識精度にどのように影響を与えるかを明らかにする必要がある。そのため文書を5つ作成し、各文書で文字間隔を-1.5, -1, 1, 2, 3, 5, 7ptと広げた文書画像35枚で実験をした。なお文字間隔が-1.5, -1ptのときは文字同士が重なっている。またフォントサイズは10.5ptで固定である。この結果を図 6に示す。

図をみると文字間隔を広くするごとに認識率が增加しているが、その後は低下している。よって文字間隔は適切な幅を選択して前処理をしなければならぬ。また一般に言われているとおり、文字が重なっていると認識率は低下するということが明らかになった。

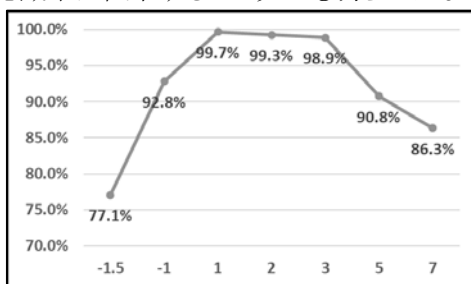


図 6 文字間隔ごとの平均認識率の変化
(縦軸: 認識率 横軸: 間隔の広さ[pt])

4. むすび

本研究では、自由手書き文字の精度低下の要因を明らかにした。その結果、文字認識のための画像化はスマホカメラで手軽に可能であることが明らかになった。また文字認識精度を向上させるには文字の大きさの均一化、文書の角度の水平化、そして行間と文字間隔を文字が重ならないように広げる前処理が効果的だと示された。

ただし、文字間隔については間隔を広げ過ぎた場合にも認識精度の低下がみられた。そのため認識精度を高めるための前処理には、適切な間隔を設定する必要がある。今回はフォントサイズを固定して実験をしたが、適切な間隔はフォントサイズに依存する可能性がある。よって文字間隔についてはさらなる実験が必要となる。

今後の課題として縦書き横書きが混在している場合の精度への影響など、他の要因が認識精度に及ぼす影響についても検討する必要がある。

参考文献

- [1] 坂井邦夫, 入江文平, 水谷博之, “知能ロボットの技術: 人工知能からのアプローチ(前編): 3.文字・文書の認識と理解” 情報処理, 44 (11), pp.1123-1129, 2003.
- [2] 黒沢由明, 入江文平, 水谷博之, 登内洋二郎, “実社会での利用が広がる文字認識技術”, 情報処理, 51 (12), pp.1530-1537, 2010.
- [3] 永田昌朗, “文字類似度と統計的言語モデルを用いた日本語文字認識誤り訂正法”, 電子情報通信学会論文誌 D, Vol.J81-D2, no.11, pp.2624-2634, 1998.
- [4] DX Suite |AI inside 株式会社, <https://inside.ai/dx-suite>, (参照 2019-2-8).
- [5] 三部祐史, 大森健児, “信頼性の低い文字認識結果に対する言語情報を用いた誤認識文字の訂正”, 情報処理, 34 (10), pp.2117-2124, 1993. <https://data.wingarc.com/ocr-and-paperless-14118/2>, (参照 2019-2-8).
- [6] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “カメラを用いた文字認識・文書画像解析の現状と課題”, 電子情報通信学会技術研究報告, TL, 思考と言語, 104 (740), pp.85-90, 2005.
- [7] 佐々木俊介, 陳奎廷, 馬場孝明, “畳み込みニューラルネットワーク(CNN)を用いた手書き日本語文字認識システムの試作”, 平成 27 年度電気・情報関係学会九州支部連合大会, 10-2A-12, p.348, 2015.
- [8] Windows10 に Python3 を入れたくなったので Anaconda でインストールしてみる[uepon 日々の備忘録, <https://uepon.hatenadiary.com/entry/2017/08/27/130515>, (参照 2018-10-30).
- [9] Python から Google Cloud Vision API を使ってみた。|極楽とんぼのロボット製作記, <https://www.g104robo.com/entry/google-cloud-vision-api>, (参照 2018-10-30).
- [10] 鈴木雅人, 北越大輔, “画像の微小スケール空間を活用した特徴点抽出に基づく手書き文字認識法に関する検討”, 信学技報, vol.117, no.432, IE2017-115, pp.257-260, 2018.
- [11] 入江文平, “文字認識技術-文字認識は終わっているか?-", 情報処理 42 (6), pp.617-620, 2001.