

文字・文書の認識・理解に関するグランドチャレンジ私案

内田 誠一†

† 九州大学 システム情報科学研究院
〒 819-0395 福岡市西区元岡 744
E-mail: †uchida@is.kyushu-u.ac.jp

あらまし 文字・文書の認識・理解問題は既にほとんど解決済のように思われるかも知れない。確かに、先人の多大なる努力により、実用化されて身近な技術となっている部分も多い。しかし実際には、現状の技術で扱える範囲においてのみ実用化されているというべきであり、工学的にもサイエンス的にも数多くの課題が残されている。本報告は、PRMU「グランドチャレンジ 2008」企画の一つの支流として、文字認識研究の未解決問題を概観した上で、そこで今後取り組むべきと思われる幾つかの課題について私案を述べたい。良きにつけ悪しきにつけ、当該分野の研究者による議論の端緒となり、また新規参入の契機を産むことを祈っている。

キーワード 文字認識, グランドチャレンジ

Challenges in character recognition research

Seiichi UCHIDA†

† Faculty of Information Science and Electrical Engineering, Kyushu University
Motooka 744, Nishi-ku, Fukuoka-shi, 819-0395 JAPAN
E-mail: †uchida@is.kyushu-u.ac.jp

Abstract Character recognition and document understanding might be seen as a solved problem, since nowadays there are many commercial and practical systems as products of great efforts by pioneers in this research area. We, however, can recognize a huge number of open problems. In this report, after reviewing those open problems briefly, several challenging problems are proposed for encouraging and inviting young researchers to this interesting and never-ending research area.

Key words grand-challenge, character recognition

1. ま え が き

文献 [1] の発刊からおおよそ 15 年経った今、電子情報通信学会パターン認識・メディア理解研究会 (PRMU) では「グランドチャレンジ 2008」と題して、今後 10 年間に当該分野で挑戦すべき課題を検討している [2]。著者はそのメンバーの一人として参加し、議論を重ねていくうちに、文字・文書の認識・理解 (以下、単に文字認識と呼ぶ) にも多くの未解決問題があることを再認識した。またそれらを再検討することが文字認識研究のみならず PRMU がカバーする研究分野全体にも有益であるとの思いを強くした。

本報告では、文字認識研究の昨今の動向および未解決問題を概観した上で、今後取り組むべきと思われる幾つかの課題について、私案を述べたい。経験・能力・知識不足から、この私案には問題も多々あると思われる。それでも、良きにつけ悪しきにつけ本報告が、パターン認識、コンピュータビジョン、メディ

ア処理、その他関連分野の研究者による議論の端緒となることを願っている。

2. 文字認識とグランドチャレンジ

各論に入る前に、文字認識それ自体が、これまでも、そしてこれからも大きなグランドチャレンジと呼ぶに相応しいことを強調しておきたい。その根拠は以下の通りである。

- 大規模な技術流を生んできた。すなわち、Tauschek 特許以来 80 年間、そこで生まれた様々な技術は他分野に転用され、逆に他分野で開発された技術が文字認識に積極的に導入されてきた。顔認識や人物認識のために、文字認識用に開発された特徴が有効であることが昨今注目されている。また音声認識で開発されたマッチングやセグメンテーション技術が、文字認識向けに転用されている。

- 多様な技術を必要とする。すなわち、セグメンテーション、前処理、特徴抽出、識別、文法制御、などのすべての要素

が成熟して初めて達成可能なチャレンジである。すべての要素にはパターン認識の本質的で困難な問題が含まれている。前処理の一部である傾き (slant) 補正ですら、これまで何百という論文が発表されながら、未解決であることはその証左であろう。

- 最も単純な形式を持ったパターンを対象としている。すなわち文字は、2 値、平面的、線分構成、サイズ小、アルファベットは有限離散、そして人工パターンである。加えて、セマンティックギャップが最も小さいパターンの一つである。これらはいずれも機械認識に適した性質である。文字認識に残された課題を鑑みすることは、一般物体認識などのより困難で自由度の高い課題を今後検討していくためにも有意義であると思われる。

- 誰にも理解しやすく、ゴールも比較的明確である。文字認識は人間の日常行為であり、従ってその問題の意味は小学生でも理解できる。また、正解が一意不変であり、達成度の定量的評価も容易である。実際これまでいくつものコンテストが開かれ、分野の技術向上に貢献している。

3. 最近 15 年間における文字認識研究の動向

1928 年、オーストリアの G. Tauschek が文字認識 (OCR) の一手法を特許出願し、それから 80 年が経った。数多くの先駆者の尽力により郵便区分機・帳票認識が実用化され、さらには極めて低価格なソフトウェアとして既に広く民間に提供されるに至った。携帯電話には名刺リーダーが搭載されているものも多い。また、文字認識を行うゲームソフトが販売されて好評を得たのは記憶に新しい。

本節では、こうした文字認識研究の歴史のうち、主に文献 [1] (1994) 以降の動向変化について概観する。改めてまとめると、検討内容の多様化が目立つ。このように、文字認識はまだ飽和していない。

3.1 ポジティブな変化

3.1.1 認識精度の向上

文字認識率は継続的に向上しているように見える。電総研 ETL-9 の認識率が 99% を超えて久しいが、その後も新しく難しい (すなわち乱雑に書かれた文字の) データベースに対するチャレンジが進んでいる。例えば、東京農工大中川研によるオンライン文字データベース (Kuchibue_d/Nakayosi_t, 1997~) の認識率も向上著しい。海外では Bern 大の Bunke らによる IAM データベース (1999~) のような、相当自由に書かれた文字列も対象とされている。また、電子辞書などに付属の手書き入力も、実際に使うとかなり高い認識率である。

データエントリーサービス企業 (例えばオーリッド株式会社 <http://www.o-rid.com/>) では、OCR の結果を海外の大量のオペレータにチェックさせることで、企業の帳票等の電子化を請け負っている。未だ何重にも人的チェックを行わなくてはならない状況は少々寂しいが、OCR にある程度以上の精度が約束されているからこそ、こうしたエントリーサービスがビジネスとして成立していると思われる。

3.1.2 情景内文字認識

文献 [1] でも既に課題として挙げられている情景内文字認識は、この 10 年で非常に活性化した分野の一つである [3]。まず



図 1 CAPTCHA

カメラの安価化、高解像度化があり、そしてユビキタス情報化が後押しした形であろうか。ケータイには、すでに名刺 OCR ソフトウェアが付属している。海外旅行者や語学学習者のために、ケータイのカメラを介したシーン中の文字認識・翻訳サービスも実現された。まさにカメラは空間スキャナである。デジカメ写真中のテキスト領域をマウスで指定すればその領域だけを OCR 処理する民生 OCR ソフトも市販されて久しい。CBDAR なる情景内文字認識に特化した国際ワークショップも今年で第 3 回目を迎える。

3.1.3 デジタルアーカイブ化

インターネット検索が一般化し、様々な雑誌・書籍類が電子出版される中、歴史的な文書のデジタルアーカイブ化の機運も高まっている。検索のためにはテキスト情報が電子化されている必要があり、ここに OCR の出番がある。文化財一般のデジタル保護の機運もあって、国内外に数多くのプロジェクトがある。紙の劣化や裏写り、インクのにじみなど、相当に難しい問題が山積している。最近では Google Book Search も登場した。対象が枯渇すると終わりのように思えるが、どのプロジェクトの話の聞いても、想像以上に埋蔵されている古文書類は多い。

3.1.4 CAPTCHA

CAPTCHA とは、図 1 のように、人間には読めるが機械には読めない文字である。インターネット上の各種サービスを受ける際、システム側が CAPTCHA を提示し、人間がそれを読んでタイプ入力することで、サービスが開始される。機械には CAPTCHA が読めないために、不正侵入ができない。これが読めれば人間と認める。言わば逆チューリングテストである。驚くほど急速に浸透し、最近では、苦勞しながら CAPTCHA を読まされる機会が多い。闇雲に読みにくくすれば人間にも読めなくなるため、文字の本質を突き詰める研究の一種と言える。

3.1.5 多言語化の波

OCR の対象となる言語が多様化している。国際会議においては、東南・南・西アジア、アフリカ言語向けの OCR 開発に関する発表がしばしば見られる。特に昨今ではアラビア文字は多い。2006 年、米国政府のサポートの下、アラビア文字と漢字 (中国の文字) に特化したワークショップ SACH'06 が開催され、文字認識分野にも軍需があるのかと噂になった。

ICDAR2007 の Arabic handwriting contest で優勝した Siemens の関係者に聞いたところによると、彼らが使った OCR エンジンには、筆記方向にあわせて Sliding window の方向を右から左にただけで、他はそのまま英文字用 OCR だったそうである。やはり文字には言語とは無関係に共通した何かがあるのかも知れない。

3.1.6 計算能力の向上

黎明期の OCR 研究は計算量制約との戦いであり、その中で

peep-hole 特徴, ゾンデ法, ペリフェラル特徴など, いわばコンパクトな特徴が見出されてきた. 現在でも計算量の少なさは依然重要であるが, 数年前の大型計算機が携帯電話サイズとなった昨今, 制約としては相当に緩和されている. 摂動法すなわち文字を様々に変形させてその都度整合を図る方法は, 非常に計算量を要する手法とされていたが, もはや十分に実行可能である. ストレージも余裕の状況である. ありとあらゆる環境中の文字をデータベース化しておき, それらを情景内文字認識に使うという, 言わば力業的な方法も始めている [4].

一般的な画像認識の分野でも, 同じく計算能力の向上を受けて, 事例・コーパスの隆盛が見られる. 文字認識は元々事例ベースで育ってきたが, その対象が拡大しつつある現在, 画像認識の様々な技術を参考にし, 文字・文書用に特化することも重要である. もちろん逆に文字認識で培われた様々な技術を, 画像認識分野に紹介する努力も忘れてはならない.

3.1.7 手書きインタフェースの民生化

手書き文字入力インタフェースは, PDA と Tablet PC により民生化が始まり, Nintendo-DS に至って (広い意味では Wii も含めて) 完全に一般的なものとなった感がある. エンタテインメントの分野で手書きが花開くとはどれぐらいの人が予想していただろう. 完全なペーパーレス化があり得ないのと同様, 書く行為自体も無くならないのであろう.

実に様々なインタフェースが開発されている中で, 最も興味深いものの一つが Anoto Pen である. これは, 細かいドットが印刷された用紙に, カメラつきペンで筆記するシステムである. ドットの動きによりペンの動き情報が得られるだけでなく, ドットのパターンそのものでどの紙面のどの位置を見ているかが同定できる. コミ Pen (株式会社フェロー/ブラザー印刷) も専用紙を使って紙上のペン先位置同定ができる意味では, 類似した目的を持つ.

3.1.8 国際会議の隆盛

文字認識関連の国際会議は以前にも増して盛況である. 当該分野最大の国際会議 ICDAR では, 2007 年に過去最高の投稿数 378 件をマークした. ICDAR2009 ではさらに増えて, 470 件との噂である. 2008 年文書画像解析システムに関する国際ワークショップ DAS が奈良で開催されたが, こちらも前回の 1.5 倍の投稿があり, 関係者を驚かせた. 手書き認識に関する国際ワークショップ IWFHR も 2008 年開催から国際会議に格上げして ICFHR となった.

こうした国際会議の隆盛には, 基礎研究よりも応用研究の多様化が寄与しているように見える. 例えば, 多言語 OCR の開発, 情景内文字・文書認識理解 (ビデオを含む), Forensic 応用, Web 文書処理, 古文書解析などは昨今活発化しているトピックであるが, いずれも応用研究の色が強い. 文字認識そのものが最初から応用指向研究であるから当然と言えば当然であるが, その黎明期において識別理論や特徴抽出理論など基礎的技術を産む分野であった事実を思うと, 多少寂しい感もある. その一方で, 情景内文字認識や紙面形状の変形補正など, コンピュータビジョンへの接近も見られる.



図 2 ドットノイズは文字を隠さないが, 線状のノイズは文字を隠す [1]



図 3 飾り文字の例

3.2 ネガティブな変化

OCR は既に無料もしくは千円-2千円で買えるソフトとして市場に流布されており, その結果, 文字文書認識は既に完成した技術と捉えられてしまう. 後述のように未解決問題が放置されている状況を見ると, 研究開発者にも「できるところまでやった」「やりつくした」といった印象もあるのではないかと. 15 年前に既にこれと似たことが言われており [1], 現在ではなおさら停滞感が強くなっている可能性もある. こうした印象が文字認識への公的・私的資金投入を難しくしているならば, それは大きな問題である.

RFID, QR コードなど, 100%に極めて近い精度で機械可読なメディアも台頭してきた. これらは文字と同様, 非接触で情報伝達が可能である. QR コードに対する違和感も薄らいできた. 実際, 携帯電話によっては, 名刺 OCR は付いていなくても, QR コードリーダは付いている. 文字には人間可読性と手書き生成可能性という強力な武器があるものの, そのテリトリーは常に新しいメディアに侵されつつもあると言える.

4. 未解決問題

応用面では様々な進捗が見られる一方で, サイエンス的にも工学的にも数多くの課題が残っていると見てよい状況である. 文献 [1] に列挙された文字認識関連の課題を見直すと, 多くは未解決のままである (注1). 手付かずの感すらある. 2008 年 2 月の PRMU (文字テーマ) では文献 [1] の執筆者の方々 4 名にご講演頂いたが, 同文献に挙げた課題が未解決である旨のご指摘が散見された.

「文字はどのような雑音や変形に強いのか」(図 2)、「“A” の様々な飾り文字を “A” と認識するメカニズムは何か」(図 3) といった, サイエンス的な課題に関する取り組みが少ないように思える. 認識とペアで語るべき文字生成 (特に活字の生成) に関

(注1): 屁理屈を承知で言えば, 文字認識は永遠に解けない問題である. ある筆者が “7” と書いて “これは “2” である」と主張した場合, 正解は一体何であろうか. このような例外は幾らでも考え付くので, 従って文字認識は完全には解けない. しかしもちろん文字認識研究は無意味ではない. 他の工学的問題と同様, 例外の影響を極力抑えながら平均的性能を最大化する努力が重要となる.

する検討の少なさも目立つ．もちろん全く無いわけではない．文字の可読性を保ったまま変形を最大化する試み (CAPTCHA) や、飾り文字をある程度認識できる技術は開発されつつあり、今後の発展が期待される．人間の視覚における選択的注意のモデル化 [5] も正攻法として重要であろう．それでも、絶対数としては少ないように思える．

工学的な課題についても、未解決課題は多い．手書き文字に対する認識率も改善の余地は多い．例えば手書きノートの内容を正確に認識するのはまだこれからの課題である．認識の方法論についても、音声認識における HMM のような定番と呼べる手法の発見には至っていない．SVM や Boosting などの機械学習法が利用されているが、これらは文字独特の性質を反映した認識技術ではない．文字の性質を反映させやすい構造解析的な手法の開発もさほど活発ではない．

5. 文字認識の課題案

以上の議論を踏まえ、今後の文字認識研究において考える課題を幾つか示す．いずれの課題も文字認識だけでなく、一般的なパターン認識問題にも関連しうるものと考えている．文献 [1] 中の未解決課題とのオーバーラップも多少ある一方で、新たに案出した課題もある．アカデミック指向や一部のトピックへの偏りについては、筆者の経験不足に因るものでご容赦いただきたい．

文献 [1] で挙げられている課題、例えば文字変形モデル [6]、視覚心理 [7]、人工文字認識 [8]、確実な棄却 [9] については、重要かつ未解決ながら以下では触れていない．これらについては、文献 [1] およびその追補版とも言える 2008 年 2 月の PRMU における講演論文 [6] ~ [9] を参考にされたい．

5.1 情景内文字認識

5.1.1 文字検出

情景内の文字を認識する前には、情景内のどこに文字があるかを検出し、それを切り出す必要がある．映像中の物体検出と同様、これは非常に困難な問題の一つである．図 4 のように、情景内には物体のエッジが織り成す無数の “I”、“L” のような線パターンが存在する．我々がそれらと文字パターンとを瞬時に峻別している原理は何であろうか．

膨大な蓄積のある「認識のための特徴抽出」に比べ、「検出のための特徴抽出」はまだ検討の余地が多いと思われる．従来法では、エッジの複雑さ、周囲とのコントラストなどを手がかりにしたものが多かった．KLT tracker で用いられる局所コーナー点検出や SIFT の際の特徴点検出はどの程度役に立つだろうか．顔検出における Haar-like 特徴のような切り札的なものは無いのだろうか．

文字デザインは、検出の困難性に影響するように思える．色調のコントラストだけでなく、字形も影響すると予想される．図 5 は検出精度の文字フォント依存性を示す実験結果例である [10]．文字が日常的にカメラに晒されるユビキタス時代 [9], [11] のために、検出容易な文字をデザインすることはできないだろうか．山本 [12] は文字を脳に淘汰されたパターンと呼んでいる．数多くのフォントから検出容易なものに淘汰できないだろうか．



図 4 環境画像．そのエッジ部には、“I”、“L”、“T”、“Y”、“M” など様々な文字形状を見ることができる．



図 5 文字形状による検出しやすさの違い [10]

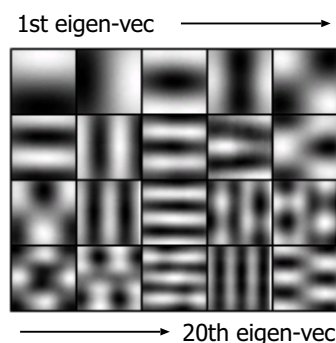


図 6 非文字画像集合の主成分

通常の文書画像の場合、切り出しの問題は文字認識と同時に解かれることが多い．すなわち文字列に対する Segmentation-by-recognition (recognition-based segmentation) である．この原理を情景画像に拡張したアルゴリズムはできないだろうか．すなわち、文字列の場合は 1 次元で済んだ切り出しを、2 次元に拡張することはできないだろうか．単純拡張では、計算量爆発の問題が発生するが、情景画像にあっても文字列自体は本質的に 1 次元であることを活かさないであろうか．例えば傾きを許した X-Y cut 法のようなアプローチはありえないだろうか．

5.1.2 文字と非文字

何らかの学習に基づいて文字検出を行う場合、負例すなわち非文字をどのように定義するのが問題になる．単に自然画像を沢山集めてくればよいように思われるが、自然画像の集合、すなわち近傍で相関の高いマルコフ的な画像の集合は、離散コサイン変換の上位基底で張られる部分空間 (DCT 低周波成分) に偏って存在するため、実情に即しているとはいづらい．図 6



図7 OCR フォントと MICR フォント

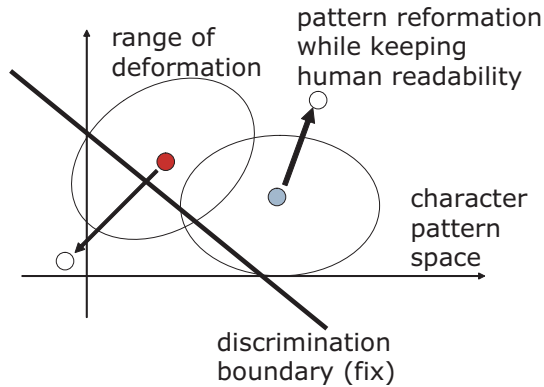


図8 機械可読な文字合成

は Caltech Background dataset から得た非文字画像集合の上位基底 (主成分) であり, DCT 低周波成分に類したものとなっている. この非文字部分空間の単純な補空間を文字空間としても限界がある. 文字の定義が困難であると同様 [1], 合理的な非文字の定義は相当に難しいと思われる. むしろ経験主義的に, 適当な画像データベースを非文字集合として定義をしたほうが, ETL-DB の貢献のように道を近くする可能性がある.

5.1.3 コンテキスト

切り出しにおいてコンテキストの積極的な利用が考えられる. PhotoBook [13] のように, 情景画像を森/空/建物などに領域分割する技術が多く検討されている. その分割結果を用いて, 文字領域の存在しない森や空を無視して, 建物付近を集中的に探索すればよい.

その逆も考えられる. すなわち, 環境中の文字を認識できれば, その環境に関する情報がある程度得られる. これは正に我々人間が日々やっていることであるが, 例えば “お会計” と書いてあれば, そこが何かの店であり, レジが近くにあることが理解できよう. さらに, 文字の大きさによる距離感, 文字列配置による被印字物体の形状や姿勢推定, 重力方向なども推定できよう. 例えば, “田” の字の各ストロークが直線的に写っていれば, 印字面は曲面ではなく, またストロークの平行関係を利用すれば, 印字面の射影変換推定が可能である. こうした環境と文字の協調認識は今後の重要な課題となると思われる.

5.2 文字合成

文字はそもそも人工的なパターンであるから, 従って機械認識に適したデザインを施すことも可能である. 1960 年代に開発された OCR・MICR フォント (図 7) がその例である. 機械可読性の高い文字フォントはコピキタス時代において人間と機械を結びつける重要なメディアとなりうる. こうしたフォント

をシステムティックに合成する方法は無いだろうか.

合成の際には, 人間可読性を保つことを忘れてはならない. 単に機械可読性が高いだけであれば, バーコードのような人間にとっては無意味なパターンが合成されるだろう. こうした状況を避け, 人間可読性を維持するためには, 例えば心理的距離 [14] のような許容範囲を規定する道具が必要であろう. また必要に応じてフォント専門家の意見も乞うべきと思われる.

さらに, 合成の際には, 情景内文字・文書に起こる様々な変動も考慮すべきである. 以上をまとめると, 図 8 のように, 与えられた識別器と変動モデルの下で, 確実に認識されるフォントを合成する問題に帰着する. この場合も, 「文字は脳に淘汰されたパターンである」という仮説 [12] は, 様々な示唆を与える.

従来も文字合成に関する検討は多数行われており, それらは主に手書き風フォント生成や学習パターン合成を目的していた. 上述の機械可読フォントに加え, 他に「機能性フォント」を考へることはできないだろうか. トナー量の使用が少なくて済むエコフォント (株式会社ニス) や DataGlyph (Xerox) などもある. 文字およびその配列の美しさが読後の作業効率に影響を及ぼすという報告 [15] も機能性フォントの参考になる.

5.3 特徴抽出

認識のための特徴抽出について膨大な検討が為されてきたことは周知の通りである. 輝度特徴, ぼけ特徴, 背景特徴, 位相特徴, 方向特徴, 射影特徴, ペリフェラル特徴 (交差特徴), フーリエ記述子, モーメント特徴など, 枚挙に暇が無い. このためか最近では, 新しい特徴についてあまり提案されていないように思われる. 「特徴抽出に王道なし」とは広く知られた言葉である. 最近のパターン認識の道具を使って, 特徴抽出について再考すべき時が来ているのではないだろうか.

文字認識で広く用いられている局所方向ヒストグラム特徴は, ある範囲の複数の特徴を集積するという点において, 一般物体認識において有効とされる Bag-of-features と関連している. また同じく有効とされる HOG 特徴は, 局所方向ヒストグラム特徴とほぼ同じものである. ならば, 一般物体認識と文字認識は全く同じ手法で解決できるのであるだろうか. できる / できないとすればその理由は何であろうか.

特徴の使用法についても, まだ再考の余地があるように思われる. 一般には 1 文字全体を同じ特徴で均質に表現することが多い. これに対し, 物理的意味の異なる様々な特徴を, 文字の部分毎もしくはカテゴリ毎に選択的に使うことも考えられる.

5.4 データ構造

今後は, 情景内文字や手書き文字を膨大なコーパスとして蓄積して利用することが考えられる. この場合, 高速な最近傍探索が必要となる. すなわち, 画像レベル, 単文字レベル, 単語レベルそれぞれで, 高速な絞込みが可能なデータ構造 (辞書構造) の実現が期待される.

高速探索を可能とするデータ構造は映像処理や物体認識の分野で近年多数提案されており, それらに文字独特の性質を組み込むことは一考の余地がある. 例えば, 文字の部首構造を活かした構造化字体表現 [16] に基づく高速最近傍探索や, 文字を対象とした主成分木・LSH の構築などはすぐに思いつく. 混合類似

度をデータ工学的視点から吟味するというのも面白いかも知れない。

kd-treeのような階層型探索において、階層の深さと認識しやすさ(読みやすさ)に相関があるようなデータ構造をシステムティックに準備できないだろうか? 文献[1]で指摘されているように、よく見るフォント(Univers67)の“A”は一瞬のうちに認識できるのに対し、派手な花文字フォント(Masquerade)の“A”では「これは“A”か?」という思考を要する。このような応答時間の差異を、ある特徴空間上でのkd-treeにおける探索時間の差異(すなわち階層の差異)として表現できれば、その時の木構造と特徴空間はある意味で人間の脳内辞書を模していると言える。

5.5 他のメディアとの共存

RFIDやQRコードなどの新しいメディアと文字の相補的利用が考えられる。すなわち、QRコードの機械可読性と文字の人間可読性を相補的に活かせる応用は無いであろうか? 例えば、QRコードから得られる少量の情報をパリティのように使って、ページ全体の文字認識誤りを自動訂正する方式が考えられる[17]。また、書籍に貼付されたRFIDによりその本のフォント情報やレイアウト情報が提供されるとすれば、OCR処理に大きく寄与するであろう。これらとは逆に、混在して受信された複数のRFID情報の区別に、RFID付近に印刷された文字情報が利用できる場合もあろう。

5.6 センサとしての人間の利用

DFKIのBuscherら[18]は、人間の視線が文書画像中のどこに注がれているかをを用い、注がれている状況が長いほどそこが重要な領域であるとして、文書分類器の学習に利用している。これはある意味、人間を文書画像の特徴抽出器として利用しようという意欲的な試みである。もちろん、人間が手で領域指定を行うような試みは以前からもあったと思われるが、この試みではそれが自動化され、人間がまさにセンサとして利用されている点が興味深い。

夢のような話かも知れないが、Brain-Computer Interface(BCI)研究が進み、ノイズの多い脳波の中から文字に反応するような信号を見出すことができれば、それを視線検出と組み合わせることで情景画像中の文字検出を実現できるかもしれない。さらに、脳波により文字カテゴリまで峻別できるとすれば、各人が日々読んでいる(そして思考している)全てのテキスト情報を自動的にログ化するという究極のOCRを実現できる。この場合の人間は、スキャナであり、文字検出・切出しセンサであり、文字認識ソフトでもある。

6. まとめ

パターン認識研究とは無関係な人々に文字認識の話をする、文字認識が未解決という事実に驚かれることがある。多くの人々の中では文字認識は既決問題、もしくは解決されて当たり前の問題とされているようである。もしかしたらパターン認識研究者の中にすら、そのように考える人々がいる可能性もある。それが誤謬であることは、本稿において強調してきたとおりである。

まだ文字認識は終わっていない^(注2)。文献[1]の「それでは、文書・文字の認識や理解に関する問題は解決済みなのであるか? いや、決してそうではない」から15年弱が経った現在も、(残念ながら)全く同じ状況である。拙稿が文字認識研究者による議論の端緒となり、また新規参入の契機を産むことを祈っている。

謝辞: 本稿は、様々な方々との議論が基となっている。過去にPRMU文字テーマで特別講演下さった方々、PRMUグランドチャレンジのメンバ、ユニバーサルパターンプロジェクトのメンバ、坂野鋭博士(NTT)には特に感謝申し上げたい。本稿には多々不備あると思われるが、それは偏に著者の理解不足、不勉強に因るものであることを申し添える。

文 献

- [1] 小川英光編著, パターン認識・理解の新たな展開—挑戦すべき課題—, 電子情報通信学会, 1994.
- [2] 鷲見和彦, 日浦慎作, 福井和広, 内田誠一, 佐藤洋一, 佐藤真一, “PRMUグランドチャレンジ2008 - 今後10年間に挑戦すべき課題を探る -,” 画像の認識理解シンポジウム, GC, p. 3, 2008.
- [3] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “デジタルカメラによる文字・文書の認識・理解,” 信学誌, vol. 89, no. 9, pp. 836-841, 2006.
- [4] 草地良規, 伊藤直己, 鈴木章, 荒川賢一, “画像インデクシングを目的としたテキスト領域抽出不要の景観中文字認識,” 信学技報, PRMU2004-89, 2004.
- [5] 福島邦彦, 大串健吾, 斎藤秀昭, 視聴覚情報処理, 森北出版, 2001.
- [6] 若原徹, “「決定論的文字変形モデル」再考,” 信学技報, PRMU2007-223, 2008.
- [7] 内藤誠一郎, “視覚情報処理の物理数学,” 信学技報, PRMU2007-224, 2008.
- [8] 山田博三, “文字を巡るアナログとデジタル,” 信学技報, PRMU2007-225, 2008.
- [9] 津雲淳, “文字認識研究の新たな展開に向けて,” 信学技報, PRMU2007-226, 2008.
- [10] 服部亮史, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, “検出容易な文字パターン生成に関する検討,” 信学技報, PRMU2008-2009. (掲載予定)
- [11] 山田敬嗣, “コピキタス情報インタフェースのための文字・文書メディア認識・理解,” 信学技報 PRMU2003-229, 2004.
- [12] 山本和彦, “脳に淘汰された文字とパターン認識技術,” 信学技報 PRMU2004-226, 2005.
- [13] R. W. Picard and T. P. Minka, “Vision texture for annotation,” Multimedia Systems, vol.3, pp.3-14, 1995.
- [14] 鶴岡信治, 村瀬晶彦, 木村文隆, 横井茂樹, 三宅康二, “人間の字種識別基準を用いた自由手書き片仮名文字認識,” 信学論, vol. J68-D, no. 4, pp. 781-788, 1985.
- [15] K. Larson, “The technology of text,” IEEE Spectrum, vol. 44, no. 5, pp. 26-31, 2007.
- [16] 末代誠仁, 中川正樹, “オンライン手書き日本語文字認識における構造化字体表現に対するプロトタイプ学習,” 信学論, vol. J86-D2, no. 1, pp. 1-11, 2003.
- [17] 岩村雅一, 内田誠一, 大町真一郎, 黄瀬浩一, “付加情報を用いるパターン認識,” 信学論, vol. J90-D, no. 2, pp. 460-470, 2007.
- [18] Georg Buscher and Andreas Dengel, “Attention-Based Document Classifier Learning,” The Eighth IAPR International Workshop on Document Analysis Systems, pp. 87-94, 2008.

(注2): MIRU2008において、グランドチャレンジテーマに関するアンケートを取ったところ、力強く「文字認識は終わっていない!」と書かれたものがあり、妙に感動した。