

認識形入力方式に関する調査研究報告書

2023年3月

一般社団法人 電子情報技術産業協会
認識形入力方式標準化専門委員会

序 文

我が国の将来の労働力人口減少に対応するためには労働生産性の向上が急務であり、デジタル技術の利活用による効率的な新しい社会づくりを目指すDX（Digital Transformation）や、業務の自動化を進めるRPA（Robotic Process Automation）といった概念に大きな注目が集まっている。当委員会の名称ともなっている認識形入力方式、すなわち文字や画像などの実世界に存在するいわゆる「パターン情報」をAI（人工知能）技術により自動的に認識し、デジタル情報システムへの入力情報に変換する枠組みは、実世界の様々な情報をデジタル世界に取り込むための極めて重要な基盤技術であり、まさに今後の社会変革の流れにおいても重要な役割を担うものであるといえる。

このような背景のもと、当協会では、2022年度「認識形入力方式標準化専門委員会」を設けて、認識形入力方式を支える基盤技術に関する議論と、それを活用したシステムの現状調査を行ったので、その成果をここに報告する。

本年度当委員会では昨年度に引き続き、認識形入力方式の性能劣化を引き起こす実環境における外乱要因について、現象・要因の観点から整理を行うとともに、新たな取り組みとしてこれを有用なガイドラインとしていくための施策について議論を行った。また、AI技術の動向を広く調査し、研究開発を推進するための方策について議論を行った。既に実用化が進んでいる「整備された環境におけるOCR技術」については、引き続き現状の技術と装置について調査を行った。さらに、今年度開催された学会等動向について調査を行い、技術の最新動向の把握と今後のあり方について議論を行った。

本報告書の作成にあたり、ご協力をいただいたユーザー、メーカー各位と、ご指導を賜った関係省庁、並びに本報告書の作成にあたって労を賜った委員各位に深く感謝の意を表するとともに、本報告書が各方面に広く利用され、我が国における情報化と産業の発展に寄与できれば幸いである。

2023年3月

一般社団法人 電子情報技術産業協会
認識形入力方式標準化専門委員会
委員長 高橋 勝彦

認識形入力方式標準化専門委員会名簿

(敬称略、順不同)

委員長	高橋勝彦	日本電気株式会社
副委員長	田中宏	富士通株式会社
監事	山合敏文	株式会社リコー
委員	佐藤雄隆	国立研究開発法人 産業技術総合研究所
委員	岩田健司	国立研究開発法人 産業技術総合研究所
委員	古畑彰夫	東芝デジタルソリューションズ株式会社
委員	田辺吉久	OCR エキスパート
委員	松村博	OCR、AI 技術アドバイザー
客員	栗田多喜夫	広島大学
事務局	吉田晃	一般社団法人 電子情報技術産業協会
事務局	塩川大介	一般社団法人 電子情報技術産業協会

(2023年3月31日現在)

目 次

1. はじめに	1
1.1 デジタル化に関する俯瞰的考察	1
1.2 審議過程	4
1.3 技術の現状調査	4
2. 実世界環境における認識機器の耐環境性の標準化	6
2.1 実世界環境における外乱要因	6
2.2 RPAにおけるUI操作自動化のためのOCRの外乱要因	22
2.3 実世界環境における認識機器(技術)の現状と今後の展望	26
2.4 OCRの品質保証	32
3. 認識技術の動向	35
3.1 文字認識技術の現状と今後の展望	35
3.2 文字認識・文書理解に関する国内学会の発表動向	40
3.3 パターン認識研究の最新動向	45
4. 文字認識システムの技術調査	53
4.1 OCRの現状	53
4.2 製品分類について	55
5. ペン入力文字認識と応用	77
5.1 ペン入力応用システム	77
5.2 教育向け応用	81
5.3 主な日本語オンライン文字認識エンジン	82
6. 海外動向調査	86
6.1 国際会議における動向について	86
6.2 国際会議の再編成についての議論	86
6.3 2022年度の開催状況	87
7. 今後の展望	93

1. はじめに

近年の AI 技術の非連続的ともいえる急激な進化に伴い、認識形入力方式の性能も年々目覚ましい向上を遂げている。また、スマートフォンに代表されるモバイル端末のコモディティー化や性能向上、及び高速なモバイルネットワーク通信網の整備なども相まって、様々な人が様々な用途に認識形入力方式を利用できるようになった。例えば、レシートや手書きメモをスマートフォンで撮影することで記載内容をデジタルデータに変換したり、身の回りにある動植物を撮影するだけでその種類を検索して調べたりすることも可能になりつつある。

このように認識形入力方式の活躍の場が拡大する一方で、利用環境の多様化に伴って認識精度に影響を及ぼす外乱が混入しやすくなっており、認識形入力方式が本来の性能を発揮できず、性能が劣化するリスクも高まっている。

このような背景のもと、2022 年度の認識形入力方式標準化専門委員会では、認識形入力方式を支える基盤技術に関する議論と、それを活用したシステムの現状調査を行った。特に、認識形入力方式の性能劣化を引き起こす実環境における外乱要因の整理と、これを有用なガイドラインとしていくための施策について審議を行った。また、進展が著しい認識形入力方式の最新動向について、製品及び技術の観点から調査を行った。

本章では、本委員会の今年度の活動報告を行うに当たり、まず、実世界のアナログ情報がデジタル化され、また、デジタル情報が必要に応じてアナログ化されて人間に利用される様子を俯瞰的に表現した。それにより、本委員会が特に課題として検討している文字情報のデジタル化変換の位置付けを明確化することを試みている。それに続き、本委員会の2022年度の活動状況と審議内容について述べる。

1.1 デジタル化に関する俯瞰的考察

前述のように社会のデジタル化は急速に進んでおり、会議やイベントのオンライン化が急務となるとともに、紙ベースでの書類のやり取りにも支障をきたす事態となり、電子化、押印廃止などの動きが強まっている。

本委員会で従来から扱っている OCR は、主に紙に印刷された文字情報を機械で読み取り、情報をデジタル化する技術であり、アナログ情報をデジタル化して活用する試みの先駆けと考えることができる。一方、現代においては既にデジタル世界（オンライン）のみに存在し、活用される情報も急速に増加している。このような状況下において「デジタル化」というキーワードを俯瞰的視点から見た場合に、具体的に何を意味していることになるのか、OCR や RPA の文脈を中心としながら本委員会で議論・整理を行った。（以降、議論の要点をまとめるが、本件は今後も継続して議論を行うこととしており、現時点までの整理の速報版であるという点に留意されたい）

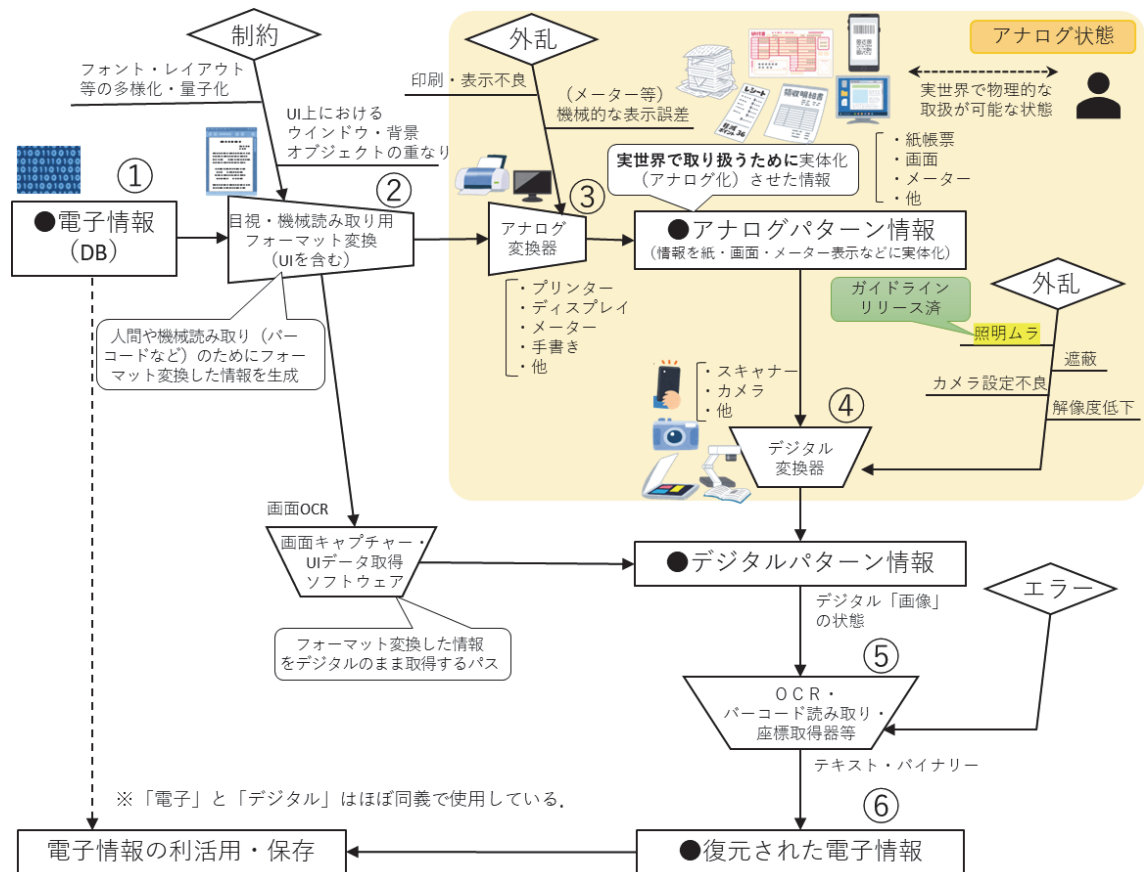


図 1.1-1 デジタル化に関する俯瞰的な考察

図 1.1-1 に考察の概略図を示す。まず、左上に①として電子情報があり、過去に電子化された情報や、初めからデジタルデータとして作成された、いわゆるボーンデジタルの情報が存在する。現代社会では、大半の情報が電子化されてインターネットや各社のオンプレミスシステム上で扱われており、電子情報は既に我々の社会に不可欠なものとなっている。

この電子情報を、図中下に向かう点線矢印のように直接利活用することが最もシンプルで効率的であると考えられる。しかしながら、自動化されたシステムを除けば、情報を活用するのは人間であり、人間が容易に読み取ることが可能になるように情報の体裁を加工することが必要となる(②)。具体的には、画面表示や紙面生成のために、文字を表現するためのフォントを割り当てたり、文章の配置をレイアウトしたりすることが考えられる。また、データを列挙するような場合には、表などとしてレイアウトすることも考えられる。体裁が加工されたデータはこの時点ではデジタル情報のままであるが、フォントやレイアウトの多様化や、表現の都合上画像化されるような場合には量子化による解像度低下の問題などが発生する可能性がある。

次に体裁を加工した情報を、人間が目視できるように何らかのアナログ情報に変換する(③)。例えば、プリンターで紙に印刷したり、ディスプレイ装置の画面に表示したりする。この段階では、印刷・表示不良などの外乱が加わる可能性がある。ディスプレイ上で情報を確認・操作するだけであれば、ディスプレイ装置以外に実世界に実体は存在しないので、デジタル情報をそのま

ま扱っているとみなすこともできる。一方、紙ベースに変換された情報は近年急速に減少しているものの、依然として大量の紙の書類や帳票類などが実世界に実体を持って我々の周りに存在している。これらをどうにか生成せずにデジタルのまま扱う、またはデジタルに戻す、ということが「デジタル化」が目指す目標の一つであると考えられる。

紙の情報をデジタル情報に戻すことを考える場合、まず、デジタル変換（④）が必要となる。これを行う代表的な装置としては、スキャナーやカメラなどが考えられるが、特に近年、スマートフォンによる帳票読み取りのようにカメラを用いるケースが増加しており、撮影時に照明ムラ、遮蔽、カメラ設定不良、解像度低下などの外乱が加わることで、後の工程に支障をきたす可能性がある（本委員会ではこのうち照明ムラに関して、ヒストグラムの分離度に基づき照明ムラのレベルをⅠ～Ⅲの3段階で定義し、それに基づき認識機器のグレード定義を行う利用ガイドラインを JEITA ITR-4010「実世界環境における OCR の利用ガイドライン 照明ムラ版」として 2019 年度にリリースしているので併せて参照されたい）。

次に、スキャナーやカメラによって得られた「デジタル画像」から文字や数値の情報を OCR などによって「デジタル情報」として抽出する（⑤）。なお、ここでも抽出は完全ではなく、ある確率でエラーが発生することに注意が必要である。

このようにしてようやく復元された情報（⑥）は、②～⑤の工程において様々な外乱を受け劣化している可能性があることに注意が必要であるが、再び電子情報として利活用・保存することが可能な状態となる。なお、近年では RPA の文脈で、画面 OCR が用いられることがある。これは本来人間に提示するためのユーザーインターフェースを機械が読み取り、人間に代わって操作・入力を行うために用いられる（もともと自動操作に対応したシステムであればこのような仕組みは必要ないが、通常の間用ソフトウェアを改修せずに、そのまま比較的容易に自動化できるということで注目されている技術である）。画面 OCR は、②で画面表示用に体裁が加工されたデータを取得し、⑤の工程に流すことで画面上の情報を読み取る。

以上の工程を振り返ってみると、明らかに元の電子情報（①）を直接活用することが最も効率的であることがわかる。しかしながら、情報化が既に進んでいるはずの近年においても紙媒体は重視され、証憑等、事実を証明するためのエビデンスとしてすら用いられてきた。これは、電子情報が存在しなかった時代から続く「従来からのやり方」を根本的に変えることが困難であったことが主な原因であると考えられるが、紙媒体は実世界で実体を持つため人間にとって扱いがわかりやすく、金庫などによる物理的手段によってある程度の保護も可能であるなどのメリットも重視されたのではないかと考えられる。

今後「デジタル化」のスローガンのもと、アナログ化を経ずに直接電子データとして扱われる情報の割合は大幅に増加するものと考えられる。しかしながら一方で、人間自身がアナログ世界に存在している以上、その利便性を考慮すると紙への印刷などアナログ化することが適切な例も

残るだろう。そのようなアナログ情報を効率よく活用するためには、エラーの影響を考慮しつつ人手を介さずにデジタルへ変換する技術が必要であり、安心して自動変換を行うための品質保証の問題も発生する。なお、今回は OCR を中心とした考察を行ったが、近年では人・場所・システムなど、実世界の様々な対象をデジタル化して扱うデジタルツインという概念も注目を浴びており、今後はこれも踏まえた議論を本委員会で行っていく予定である。

1.2 審議過程

委員会は2022年5月から2023年2月まで計8回オンライン開催され、「実世界環境における認識機器の耐環境性の標準化」について審議が行われた。

1.1節で触れたように、認識形入力方式を搭載した機器（以下、認識機器と呼ぶ）が認識対象とする情報はデジタル化の過程において多様な外乱の影響を受ける。そのため、条件によっては認識性能が劣化することが知られている。しかしながら、その条件やメカニズムは非常に複雑であるため、従来その対処は専門家の暗黙知に頼らざるを得ない面も強かった。このような知識を部分的にでも形式知化することができれば、より合理的で積極的な性能向上を図ることが検討可能になるほか、AIの学習過程にその知識を埋め込むことで認識性能を更に向上させることなども検討可能になる。

このような背景から、当委員会では、一昨年度から昨年度にかけて、OCRに影響を及ぼす多種多様な外乱を網羅的に列挙するとともに、現象・原因の観点から整理を試みた。本年度は引き続き整理を進めるとともに、整理結果をより多くの開発者やユーザーに活用してもらえるようにするための施策について議論を行った。

1.3 技術の現状調査

1.3.1 認識方式と認識技術の動向調査

DL（Deep Learning; 深層学習）等の技術革新を背景に、認識方式及びそれをを用いた装置・ソフトウェアの、性能・機能・適用対象に大きな変化が起きている。このような背景を踏まえ、文字認識・文字理解を中心とした認識方式、及びそれをを用いた装置・ソフトウェアに関して、最新動向を調査し本報告書にまとめた。また、最新動向の調査・把握のため、産業技術総合研究所の片岡 裕雄氏に「数式から自動学習するAI」、同じく産業技術総合研究所の岩田 健司委員に「安全なドローン・産業応用のための画像AI技術」とそれぞれ題したご講演を委員会中オンラインで行って頂いた。

1.3.2 文字認識関連技術の調査

カメラ付き携帯デバイスによる文字認識技術等、認識形入力方式の新たな展開が期待される応

用分野について調査を行った。また、文字認識装置についての現状を調査するとともに、OCR 製品一覧表の更新・拡充を行った。

1.3.3 技術動向調査

今年度開催された各会議のうち、今後、認識形入力方式に大きな影響を与えると予想される最新の研究成果の調査を行い、委員会において今後の変化・発展の可能性について議論を行った。また、調査結果の要約を本報告書にまとめた。

2. 実世界環境における認識機器の耐環境性の標準化

AI 技術の急激な進化に伴い、認識形入力方式を搭載した認識機器の用途が急速に拡大している。その一方で、1.1 節で述べたように利用環境下における外乱の影響を受けて認識精度が変化するという認識形入力方式の特性と、用途の拡大に伴う外乱の混入機会増大とから、条件によっては認識機器が本来の性能を発揮できず、認識精度が劣化するリスクが高まりつつある。そして、このことが認識機器の品質管理を難しくしつつある。

このような背景から、当委員会では認識機器の代表格として特に OCR に注目し、OCR の普及を促進するための議論を進めてきた。ユーザーが OCR を正しく利用できるようにすることを目的として、外乱要因の中でも認識精度への影響が大きい照明ムラに着目し、OCR の利用ガイドラインを 2020 年度 4 月に制定した。続いて、実世界環境下で認識精度に影響を及ぼす要因とそれによって生じる画像劣化の現象を要因表として整理し、今年度は特に要因表の活用を促進する施策について議論した。

本章では、まず、2.1 節にて要因表の整理状況及びその活用施策検討状況について報告する。2.2 節では OCR の新たな活用方法として注目される RPA における UI 操作自動化について説明する。2.3 節では実世界環境における認識機器の最新動向と今後の展望について述べる。2.4 節では、OCR の品質管理をめぐる最新動向について述べる。

2.1 実世界環境における外乱要因

1.1 節で述べたようなスタンド型のスキャナーやカメラによる文書・帳票画像のデジタル化の過程においては、図 2.1-1 に示すように(a)光源、(b)障害物、(c)対象物、(d)カメラ、(e)撮影者が関わり、これらに起因して外乱要因が発生すると考えられる。昨年度は、外乱要因をその発生源ごとにグループ化するとともに、外乱要因によって生じる画像劣化の現象との関係を整理し、要因表にまとめた。今年度は要因表の活用促進を目的として、主に以下の 3 点について審議を行った。

- ・活用促進を目的とした要因表の改版
- ・要因表のダイジェスト版
- ・要因表の活用方法

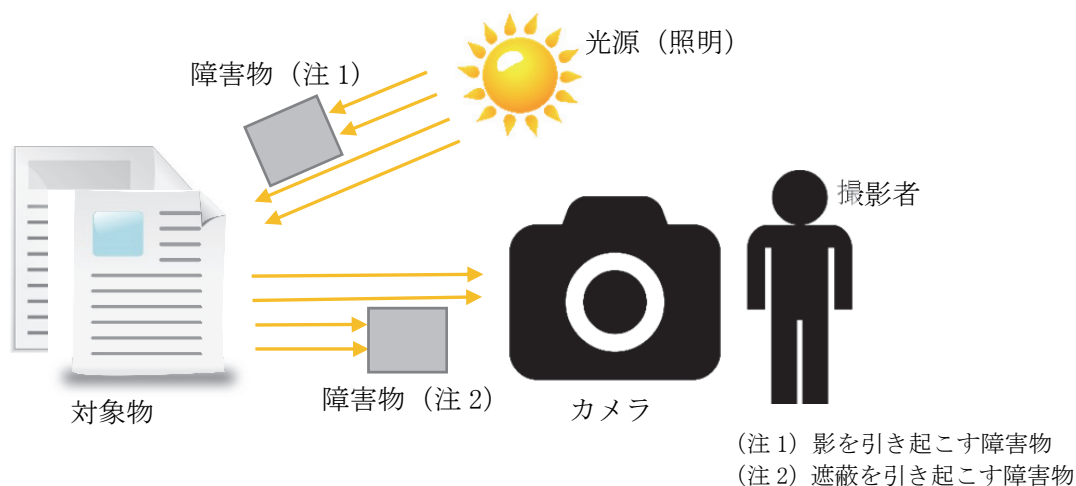


図 2.1-1 文書・帳票画像のデジタル化の過程

2.1.1 活用促進を目的とした要因表の改版

昨年度は要因の軸に着目して要因表の整理を行ったが、要因表の活用を促進する上では、軸だけでなく、各要因が認識精度に影響を与える条件も記載されている方が使い勝手がよい。そこで、認識精度に影響を与える条件が明確でなかった要因について条件を追記した。また、技術に明るくないユーザーによる理解を補助するため、OCRに影響を与える現象を示すイメージ図を追加した。改版した要因表を表 2.1-1 に示す。表中、L・xx（xx は数字）などの表記は、外乱要因とそれに起因して画像上に生じる現象との対応関係を示す。表からわかるように、1 つの外乱要因が複数の現象を生じさせる場合もある。また、要因表は認識対象物をスマートフォン内蔵カメラやスタンド型スキャナーで撮影して OCR 認識する際に認識精度に影響を及ぼす要因や現象を示すが、一部の要因については従来型のスキャナーで画像を取得して OCR 認識する場合にも当てはまる。これらを区別できるよう、表 2.1-1 では従来型のスキャナーで画像を取得した場合にも共通して当てはまる要因については青字で記載した。

表 2.1-1 OCR に影響を与える要因及び現象





分類	現象	要因
光源	<ul style="list-style-type: none"> • 黒つぶれ L-01, L-04, L-06, L-07, L-10 	<ul style="list-style-type: none"> • 状態及び種類 <ul style="list-style-type: none"> - 明るすぎる/暗すぎる (L-01) - ムラの出る照明 (L-02) - 点光源 (L-03) - 周波数 (フリッカー) (L-04) - フラッシュ (L-05) - 色がついた光 (L-06) • 位置及び数 <ul style="list-style-type: none"> - 光源から対象物までの距離が近い (L-07) - 位置が局在 (L-09) • ライティング用品 (レフ板、ルーバー、デフューザーなど) の不適切な使い方 (L-10)
	<ul style="list-style-type: none"> • 白飛び L-01, L-05, L-07, L-10 	
	<ul style="list-style-type: none"> • シェーディング L-02, L-03, L-05, L-07, L-09, L-10 	
	<ul style="list-style-type: none"> • 低コントラスト L-01, L-06, L-07, L-10 	

表 2.1-1 OCR に影響を与える要因及び現象 (続き)

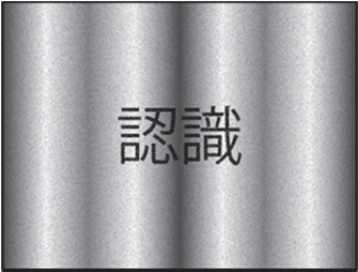


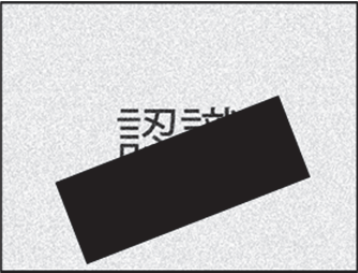
分類	現象	要因
光源	<ul style="list-style-type: none"> ・ 縞々 L-04, L-10 	
障害物	<ul style="list-style-type: none"> ・ 像がゆがむ 0-03, 0-06, 0-08 ・ 影がおきる 0-01 ・ 遮蔽される 0-02, 0-06   	<ul style="list-style-type: none"> ・ 影を引き起こす障害物 (撮影者自身 (頭、手)、カメラなど) (0-01) ・ 遮蔽を引き起こす障害物 (手の指、紙面の重なり、前に立つ人の頭など) (0-02) ・ 媒体 (水、ガラスなど) (0-03) ・ 媒体中の障害物 (霧、雨、雪など) (0-06) ・ レンズへの付着物 (皮脂、結露、コーティングなど) (0-07) ・ 対象物への付着物 (結露など) (0-08)

表 2.1-1 OCR に影響を与える要因及び現象 (続き)



分類	現象	要因
障害物	<ul style="list-style-type: none"> • ぼける 0-06, 0-07  <ul style="list-style-type: none"> • 低コントラスト 0-06 	

表 2.1-1 OCR に影響を与える要因及び現象 (続き)

(注) 青字はスキャナーベースの OCR にも共通する外乱要因を示す。

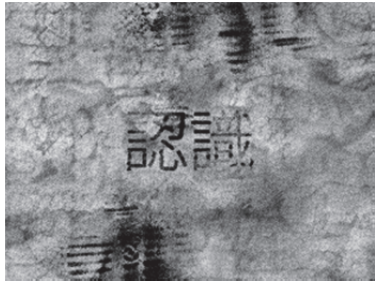
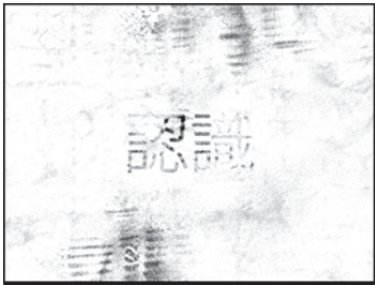

分類	現象	要因
対象物	<ul style="list-style-type: none"> • 汚れている T-07  <ul style="list-style-type: none"> • テカっている T-01, T-02, T-03, T-04, T-24, T-25 • かすれている T-13, T-16, T-17, T-25  <ul style="list-style-type: none"> • ゆがんでいる T-05, T-06, T-21, T-22, T-23 • 背景がまぎらわしい T-08, T-09, T-10, T-11, T-19, T-27, T-41, T-42, T-50, T-52, T-53  <ul style="list-style-type: none"> • 文字が読みにくい T-19, T-23, T-25, T-31, T-34, T-37, T-38, T-44, T-45, T-47, T-51, T-54 • 文字が劣化している T-02, T-04, T-14, T-15, T-16, T-17, T-18, T-20, T-49 • 文字を分離しづらい T-09, T-10, T-11, T-12, T-29, T-35, T-36, T-39, T-40, T-42, T-43, T-50 	<p>[画像・画質レベル]</p> <ul style="list-style-type: none"> • ベース <ul style="list-style-type: none"> -物性 <ul style="list-style-type: none"> -光沢のある紙 (コート紙、複写紙、光沢紙、トレーシングペーパーなど) (T-01) -モニター画面(T-02)、自発光(T-03) -プロジェクター画面(T-04) -空間構成 <ul style="list-style-type: none"> -立体(T-05) -曲面 (皺、波うち、折れ) (T-06) -汚れ・模様 <ul style="list-style-type: none"> -汚れ(T-07) -背景模様 (copy 防止文字、柄が認識対象に似ている、背景が複数色で構成され文字との分離が難しいなど) (T-08) -裏写り(T-09)、透け(T-10) • カメラ OCR における情景 (背景) <ul style="list-style-type: none"> -認識対象が背景と似ている (T-11) • 文字 <ul style="list-style-type: none"> -色が一律でない(T-12) -印字不良 <ul style="list-style-type: none"> -かすれ(T-13) -つぶれ(T-14) -虫食い文字(T-15) -滲み(T-16) -インクこすれ(T-17) -濃度が一律でない(T-18) -経年劣化 <ul style="list-style-type: none"> -インクがとび圧力跡だけ残ったもの (T-19) -感熱紙の文字劣化(T-20) -回転(T-21) -変形(T-22) -ゆがんだ手書き文字、不正確な手書き文字(T-23) -印字文字 (反射) (T-24) -インク特性 (うすい、てかる、メタリック) (T-25) -かくれ文字 (X線透過) (T-26) -ベースと文字のコントラストが小さい(T-27)

表 2.1-1 OCR に影響を与える要因及び現象 (続き)



分類	現象	要因
対象物	<ul style="list-style-type: none"> • 文字列を検出しづらい T-28, T-29, T-30, T-32, T-33, T-48  <ul style="list-style-type: none"> • 類似文字 T-46  <ul style="list-style-type: none"> • 隠れている文字 T-26 	<p>[論理レベル]</p> <ul style="list-style-type: none"> • レイアウト <ul style="list-style-type: none"> - 段落レベル <ul style="list-style-type: none"> - 縦書き・横書きの混在 (T-28) - 図の中に文字がある (T-29) - 文字列のレベル <ul style="list-style-type: none"> - 直線状に並んでいない文字 <ul style="list-style-type: none"> - ルビ (T-30) - 上付き文字、下付き文字 (T-31) - 二次元的に並んだ文字 (化学式、数式など) (T-32) - 行内でのサイズの混在 (T-33) - 字体変化 (T-34) - 文字接触 (合字、カーニング、重なり、重畳) (T-35) • 文字 <ul style="list-style-type: none"> - 白黒反転文字 (T-36) - 文字種 (サイズが小さい) (T-37) - 文字装飾 (文字飾り) (T-38) - 記入枠からはみ出し (T-39) • 非文字 <ul style="list-style-type: none"> - 罫線との重畳 (T-40) - 判子の重畳 (T-41) - 網掛け (T-42) - アンダーライン (T-43) <p>[認識対象 (カテゴリ)]</p> <ul style="list-style-type: none"> • 文字 <ul style="list-style-type: none"> - 言語<日本語記入欄、英語記入欄など> (T-44) - 文字種 <ul style="list-style-type: none"> - 一般的でない字体 (T-45) - 異体字 (T-46) - 特殊なフォント (T-47) - 化学式や数式などで使われる特殊文字 (T-48) - 文字装飾 <ul style="list-style-type: none"> - ドット文字 (T-49) - 彫り付け文字 (T-50) - ロゴ (T-51) - 立体文字 (T-52) - エンボス (T-53) - 芸術的な文字 (T-54)

表 2.1-1 OCR に影響を与える要因及び現象 (続き)









分類	現象	要因
カメラ /撮影者	<ul style="list-style-type: none"> ・ 黒つぶれ C-03, C-04, C-06, C-12, C-17, C-18, C-19, C-30 	<ul style="list-style-type: none"> ・ 光学系の問題 <ul style="list-style-type: none"> - レンズの特性 - 解像度が低い(C-01) - 広角すぎる(C-02) - レンズの開口(C-03) - 周辺劣化 <ul style="list-style-type: none"> - 減光(C-04) - 解像度低下(C-23) - 内面反射(C-07) - 被写界深度が不足(C-08) - 収差(C-09) - 歪み(C-10) (見え消し) ・ 撮像素子 <ul style="list-style-type: none"> - 画素数が少ない(C-11) - ダイナミックレンジが狭い(C-12) - スミア(C-15) - ローリングシャッター(C-16) ・ 機構系 <ul style="list-style-type: none"> - 露出が高すぎ/低すぎ (C-17) - EV 値 が高すぎ/低すぎる(C-18) - 絞りすぎ/絞らなさすぎる(C-19) - シャッター速度が速すぎる/遅すぎる(C-20) ・ 画像情報の処理 <ul style="list-style-type: none"> - 符号化圧縮ノイズ(C-21) - 高圧縮PDF化による文字背景の低コントラスト化(C-22) - 多重露光処理の失敗(C-23) - ベイヤー配列(C-24) (見え消し) - ホワイトバランス(C-25) (見え消し) ・ 撮影条件 <ul style="list-style-type: none"> - 距離が遠すぎる(C-26) - 視点が被写体に対して正対していない(C-27) - カメラの保持(手ブレ)(C-28) - カメラの設定ミス <ul style="list-style-type: none"> - ピント(C-29) - 露出(C-30) - ズーム(C-31)
	<ul style="list-style-type: none"> ・ 白飛び C-03, C-12, C-15, C-17, C-18, C-19, C-30 	
	<ul style="list-style-type: none"> ・ ピンぼけ C-03, C-08, C-09, C-29 	
	<ul style="list-style-type: none"> ・ 二重(多重)にみえる C-07, C-09, C-23, C-28 	

表 2.1-1 OCR に影響を与える要因及び現象 (続き)

分類	現象	要因
カメラ/ 撮影	<ul style="list-style-type: none"> ゆがむ C-02, C-10, C-16, C-27 	
	<ul style="list-style-type: none"> 低コントラスト C-04, C-12, C-17, C-18 	
	<ul style="list-style-type: none"> 低解像度 C-01, C-02, C-11, C-26, C-31 	
	<ul style="list-style-type: none"> ノイズ (ホワイトノイズ、縞々ノイズ、ブロックノイズ) C-15, C-20, C-21, C-24, C-25 	

なお、要因表に挙げた要因及び現象は、文字認識のフローチャートと対応付けて捉えることができる。文字認識のフローチャートの一例を図 2.1-2 に示す。この図は、文字切り出し～分類ま

での処理について、(a)DL 不使用、(b)文字切出以降の処理について DL を使用、(c)すべての処理に DL を使用、の 3 通りの実装方式をひとまとめに描いたものである。この場合、一例として、図 2.1-2 に示す処理ステップには表 2.1-2 に示すような現象が影響を与えると考えられる。

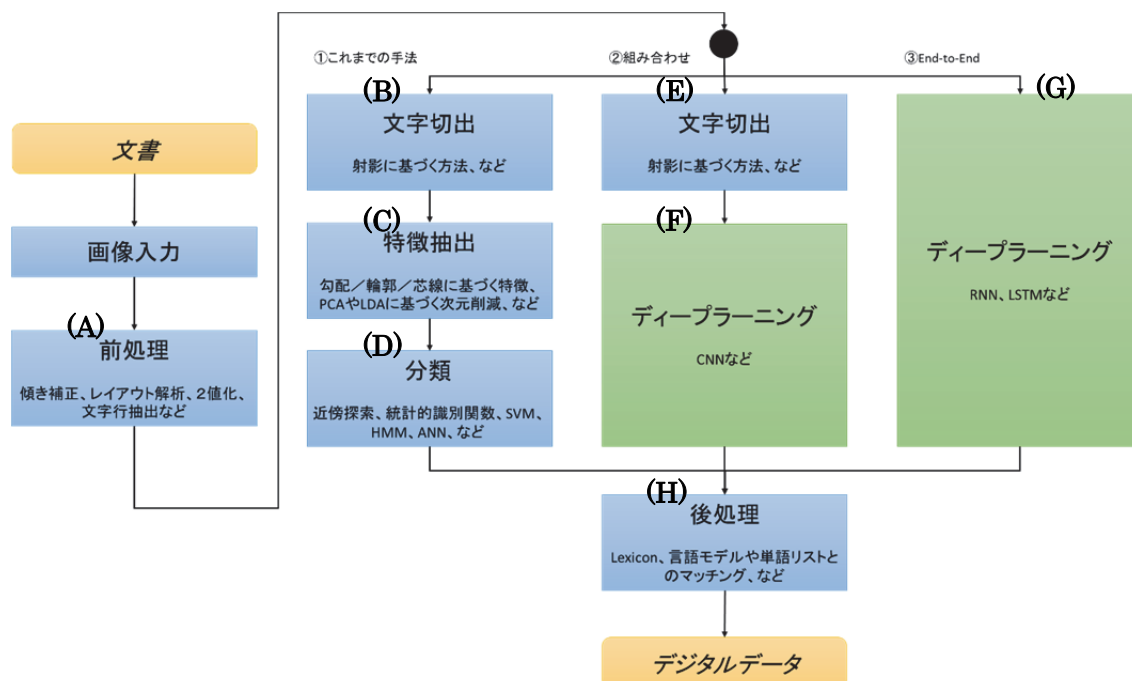


図 2.1-2 文字認識のフローチャートの例

表 2.1-2 文字認識の処理ステップに影響を与える現象の例

処理ステップ	左記処理ステップに影響を与える現象の例
(A)前処理	黒つぶれ、白飛び、シェーディング、低コントラスト、影がおきる、遮蔽される、テカっている、かすれている、文字列を検出しづらい
(B)文字切出	文字を分離しづらい
(C)特徴抽出	黒つぶれ、白飛び、ぼける、文字が劣化している、ピンぼけ、二重（多重）にみえる、低解像度、ノイズ
(D)分類	像がゆがむ、文字が読みにくい、類似文字
(E)文字切出	文字を分離しづらい

2.1.2 要因表のダイジェスト版

前節で示した要因表は、主として認識形入力方式の専門家向けに、OCR に影響を与える要因及び現象をできるだけ網羅するように作成されている。そのため、通常は起こりにくい要因・現象も含んでおり、技術に明るくないユーザーにとっては難解で扱いづらい。そこで、主に技術に明るくないユーザーに向けて、カメラで撮影した画像に対して OCR を適用する典型的なユースケースとして、オフィスの机の上に名刺・レシート・伝票などの書類を並べて、スマートフォンで撮影・OCR 認識するケースを想定し、その場合の主要な要因・現象のみを抜粋した要因表のダイジェ

スト版を作成した。本ダイジェスト版を表 2.1-3 に示す。

表 2.1-3 OCR に影響を与える要因及び現象

分類	現象	要因
光源	<ul style="list-style-type: none"> ・ 黒つぶれ L-01, L-07, L-10 ・ 白飛び L-01, L-07, L-10 ・ シェーディング L-02, L-07, L-10 ・ 低コントラスト L-01, L-07, L-10 	<ul style="list-style-type: none"> ・ 状態及び種類 <ul style="list-style-type: none"> - 明るすぎる/暗すぎる (L-01) - ムラの出る照明 (L-02) ・ 位置及び数 <ul style="list-style-type: none"> - 光源から対象物までの距離が近い (L-07) ・ ライティング用品 (レフ板、ルーバー、デフューザーなど) の不適切な使い方 (L-10)
障害物	<ul style="list-style-type: none"> ・ 影がおきる O-01 ・ 遮蔽される O-02 	<ul style="list-style-type: none"> ・ 影を引き起こす障害物 (撮影者自身 (頭、手)、カメラなど) (O-01) ・ 遮蔽を引き起こす障害物 (手の指、紙面の重なり、前に立つ人の頭など) (O-02)
対象物	<ul style="list-style-type: none"> ・ 汚れている T-07 ・ テカっている T-01 ・ かすれている T-13 ・ ゆがんでいる T-06, T-21 ・ 背景がまぎらわしい T-19, T-41, T-42, T-50, T-52, T-53 ・ 文字が読みにくい T-19, T-37, T-44, T-45, T-47, T-51, T-54 	<p>[画像・画質レベル]</p> <ul style="list-style-type: none"> ・ ベース <ul style="list-style-type: none"> - 光沢のある紙 (コート紙、複写紙、光沢紙、トレーシングペーパーなど) (T-01) - 曲面 (皺、波うち、折れ) (T-06) - 汚れ・模様 (汚れの付着、背景模様など) (T-07) ・ 文字 <ul style="list-style-type: none"> - 印字不良 (かすれ、つぶれなど) (T-13) - 経年劣化 (感熱紙の文字劣化など) (T-19) - 回転 (T-21)

表 2.1-3 OCR に影響を与える要因及び現象 (続き)

分類	現象	要因
	<ul style="list-style-type: none"> • 文字を分離しづらい T-35, T-36, T-39, T-40, T-42, T-43 • 文字列を検出しづらい T-30, T-33 • 類似文字 T-46 	<p>[論理レベル]</p> <ul style="list-style-type: none"> • レイアウト <ul style="list-style-type: none"> - 直線状に並んでいない文字 (T-30) - 文字接触 (合字、カーニング、重なり、重畳) (T-35) • 文字 <ul style="list-style-type: none"> - 白黒反転文字 (T-36) - 文字種 (サイズが小さい) (T-37) - 記入枠からはみ出し (T-39) • 非文字 <ul style="list-style-type: none"> - 罫線との重畳 (T-40) - 判子の重畳 (T-41) - 網掛け (T-42) - アンダーライン (T-43) <p>[認識対象 (カテゴリ)]</p> <ul style="list-style-type: none"> • 文字 <ul style="list-style-type: none"> - 言語<日本語記入欄、英語記入欄など> (T-44) - 文字種 <ul style="list-style-type: none"> - 一般的でない字体 (T-45) - 異体字 (T-46) - 特殊なフォント (T-47) - ロゴ (T-51)
カメラ/ 撮影者	<ul style="list-style-type: none"> • ピンぼけ C-29 • 二重 (多重) にみえる C-28 • ゆがむ C-27 • 低解像度 C-26 • ノイズ (ホワイトノイズ、縞々ノイズ、ブロックノイズ) C-21 	<ul style="list-style-type: none"> • 画像情報の処理 <ul style="list-style-type: none"> - 符号化圧縮ノイズ (C-21) • 撮影条件 <ul style="list-style-type: none"> - 距離が遠すぎる (C-26) - 視点が被写体に対して正対していない (C-27) - カメラの保持 (手ブレ) (C-28) - カメラの設定ミス ピント (C-29)

2.1.3 要因表の活用方法

上記の要因表は様々な用途に活用可能であり、例えば、以下のような使い方が考えられる。

- (a) OCR 技術者がある用途向けに OCR を開発する際に、用意すべき学習データのバリエーション (オーグメンテーション) の軸を知る
- (b) ユーザーが OCR をある用途に利用したところ認識精度が十分でなかった場合、撮影した画像の状態から再確認すべき要因を知る

- (c) ユーザーがある用途に OCR を利用しようとする際に、その利用環境に基づいて、考えた方がよい要因を知る

本節では要因表の活用のヒントを提供するために、(a)、(b)の例をケーススタディとして取り上げ、活用方法を具体的に説明する。

- (a) 活用例 1 : 技術者がある用途向けに OCR を開発する際に、用意すべき学習データのバリエーション（オーグメンテーション）の軸を知る。

例えば、図 2.1-3 に示す手順に従って、要因表から用意すべき学習データのバリエーションを絞り込むことができる。以下、一例として、スーパーのチラシに印刷された金額を読み取る OCR を開発するケースを想定して、各ステップの内容について詳しく説明する。

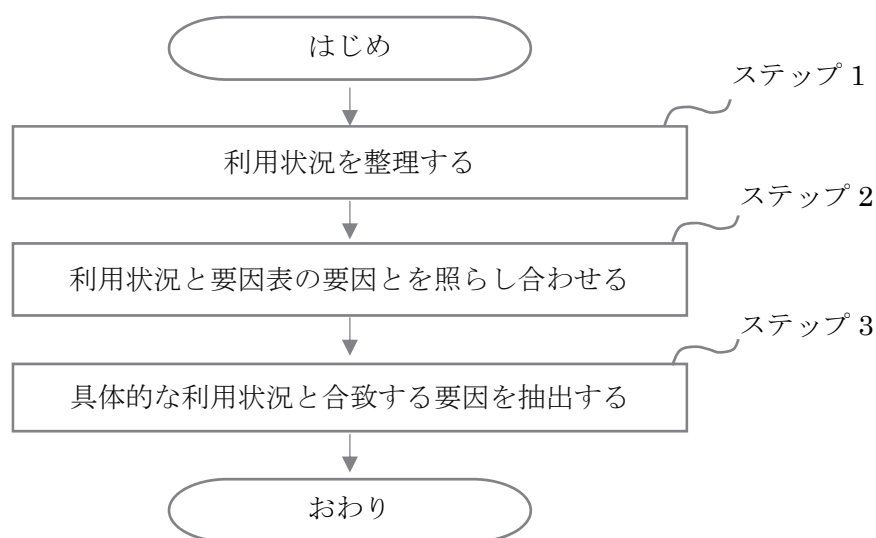


図 2.1-3 用意すべき学習データのバリエーションを知る際のフローチャートの例

- ・ステップ 1 : 利用状況を整理する

まず、開発する OCR の利用状況を整理する。スーパーのチラシに印刷された金額を読み取る OCR を開発するといった場合であれば、例えば、利用状況として、

- ・ チラシを事務所の机の上になるべく平らに広げる
- ・ 机の脇に立ってスマートフォンを手にとってチラシの画像を撮影する

などを整理する。

- ・ステップ 2 : 利用状況と要因表の要因とを照らし合わせる

ステップ 1 で整理した利用状況と要因表の要因とを照らし合わせ、それらが合致するか否か検討する。例えば、障害物に関する要因について、表 2.1-4 のように整理することができる。光源、対象物、カメラ/撮影者についても同様に利用状況と要因表を照らし合わせて整理する。

表 2.1-4 チラシ読み取りを想定したケースでの、利用状況と要因表の要因との照合例

要因	要因No.	具体的な利用状況	要因と利用状況が合致するか否か
・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）	0-01	撮影者自身などによる影が生じる可能性がある	一致
・遮蔽を引き起こす障害物（手の指、紙面の重なり、前に立つ人の頭など）	0-02	ストラップなどによる隠蔽が生じる可能性がある	一致
・媒体（水、ガラスなど）	0-03	指向性がほぼない LED 照明である	不一致
・媒体中の障害物（霧、雨、雪など）	0-06	屋内なので天候の影響はない	不一致
・レンズへの付着物（皮脂、結露、コーティングなど）	0-07	レンズはきれいに使用して使用する	不一致
・対象物への付着物（結露など）	0-08	チラシに付着する物はない	不一致

- ・ステップ3：具体的な利用状況と合致する要因を抽出する

ステップ2の整理から、利用状況が合致する要因のみを抽出する。抽出された要因が用意すべき学習データのバリエーションの軸を示す。例えば、表 2.1-4 からは、影を引き起こす障害物、遮蔽を引き起こす障害物、についてバリエーションを揃えればよいと知ることができる。

表 2.1-5 利用状況と合致する要因を抽出した場合の一例

要因	要因No.	具体的な利用状況	要因と利用状況が合致するか否か
・影を引き起こす障害物（撮影者自身（頭、手）、カメラなど）	0-01	撮影者自身などによる影が生じる可能性がある	一致
・遮蔽を引き起こす障害物（手の指、紙面の重なり、前に立つ人の頭など）	0-02	ストラップなどによる隠蔽が生じる可能性がある	一致

- (b) 活用例2：ユーザーが OCR を利用した時に認識精度が十分でなかった場合、撮影した画像の状態から再確認すべき要因を知る。

例えば、図 2.1-4 に示す手順に従って、撮影時の注意点を絞り込むことができる。

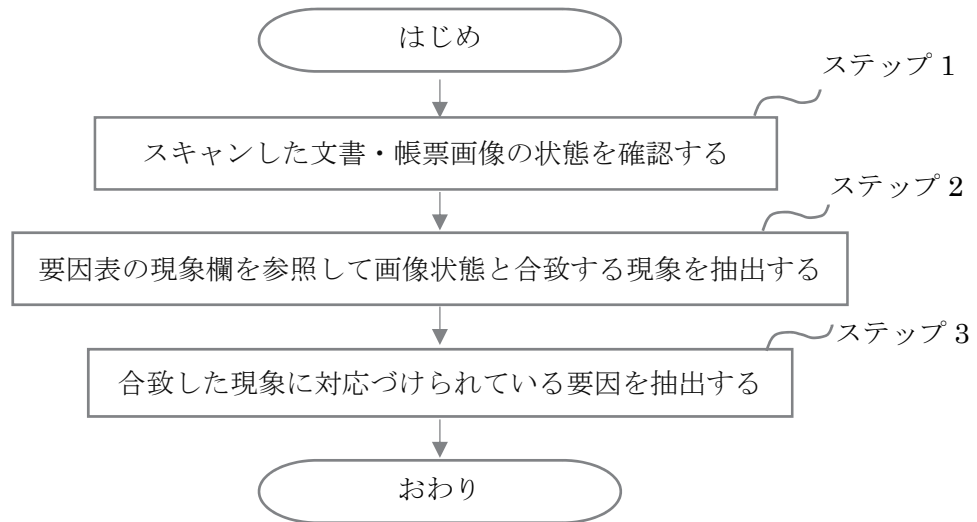


図 2.1-4 スキャンした文書・帳票画像の状態から撮影時の注意点を知る際のフローチャートの例

・ステップ 1：スキャンした文書・帳票画像の状態を確認する

スキャンした文書・帳票画像の状態（画像中に見られる劣化の現象）を確認する。例えば、画像状態として、

- 白飛びしている
- 照明ムラ（シェーディング）がある
- 対象物にテカリがある

といった項目を抽出する。

・ステップ 2：要因表の現象欄を参照して、画像状態と合致する現象を抽出する

要因表の現象欄を参照して、ステップ 1 で抽出した画像状態と合致する現象を抽出する。上記画像状態を想定した場合の抽出結果の例を表 2.1-6 に示す。

表 2.1-6 画像状態に合致する現象を抽出した例

分類	現象
光源	<ul style="list-style-type: none"> ・ 白飛び L-01, L-05, L-07, L-10 ・ シェーディング L-02, L-03, L-05, L-07, L-09, L-10
障害物 対象物	<ul style="list-style-type: none"> ・ テカっている T-01, T-02, T-03, T-04, T-24, T-25
カメラ/ 撮影者	<ul style="list-style-type: none"> ・ 白飛び C-03, C-12, C-15, C-17, C-18, C-19, C-30

- ・ステップ3：合致した現象に対応づけられている要因を抽出する

要因表を参照して、ステップ2で抽出した現象に対応づけられている要因を抽出する。表 2.1-6 に示した例に対しては、表 2.1-7 に記載した要因が抽出される。ここで抽出された要因が撮影時の注意点を示す。この中に現在の OCR の使用条件に合致する要因が含まれる場合は、当該要因が合致しないようにできないか検討する。例えば、撮影時にフラッシュをたいていたならば、フラッシュを使用せず撮影を行うようにする。

表 2.1-7 ステップ2で抽出した現象に対応づけられている要因を抽出した例

分類	現象	要因
光源	<ul style="list-style-type: none"> ・白飛び L-01, L-05, L-07, L-10 ・シェーディング L-02, L-03, L-05, L-07, L-09, L-10 	<ul style="list-style-type: none"> ・状態及び種類 <ul style="list-style-type: none"> -明るすぎる/暗すぎる (L-01) -ムラの出る照明 (L-02) -点光源 (L-03) -フラッシュ (L-05) ・位置及び数 <ul style="list-style-type: none"> -光源から対象物までの距離が近い (L-07) -複数の位置関係位置が局在 (L-09) ・ライティング用品 (レフ板、ルーバー、デフューザーなど) の不適切な使い方 (L-10)
対象物	<ul style="list-style-type: none"> ・テカっている T-01, T-02, T-03, T-04, T-24, T-25 	<ul style="list-style-type: none"> ・ベース <ul style="list-style-type: none"> -物性 <ul style="list-style-type: none"> - 光沢のある紙 (コート紙、複写紙、光沢紙、トレーシングペーパーなど) (T-01) - モニター画面 (T-02)、自発光 (T-03) - プロジェクター画面 (T-04) ・文字 <ul style="list-style-type: none"> -印字文字 (反射) (T-24) - インク特性 (うすい、てかる、メタリック) (T-25)
カメラ / 撮影者	<ul style="list-style-type: none"> ・白飛び C-03, C-12, C-15, C-17, C-18, C-19, C-30 	<ul style="list-style-type: none"> ・光学系の問題 <ul style="list-style-type: none"> -レンズの特性 <ul style="list-style-type: none"> - レンズの開口 (C-03) ・撮像素子 <ul style="list-style-type: none"> - ダイナミックレンジが狭い (C-12) - スミア (C-15) ・機構系 <ul style="list-style-type: none"> -露出が高すぎ/低すぎる (C-17) -EV 値 が高すぎ/低すぎる (C-18) -絞りすぎ/絞らなさすぎる (C-19) ・撮影条件 <ul style="list-style-type: none"> -カメラの設定ミス <ul style="list-style-type: none"> - 露出 (C-30)

2.1.4 今後の予定

今後は、多くの OCR ユーザーや開発者に要因表を活用いただけるよう、要因表及び活用方法の表現の見直しを進める予定である。

2.2 RPA における UI 操作自動化のための OCR の外乱要因

RPA とは、従来人手で行っていたコンピューター操作（UI 操作）を自動実行する技術である。人が目視で行っていた処理を自動化するために OCR が用いられるので、ここでは RPA において OCR に影響を与える要因について記述する。

RPA では、OCR は 2 種類の方法で利用される。

- (1) 紙帳票を OCR で読み込んでデータ化し、そのデータに基づいて RPA が自動処理を実行する。【帳票 OCR】
- (2) 画面上の UI 操作を自動化（自動運転）することを目的として、操作座標を求めるため、または画面に表示されているテキストをデータとして取得するために、PC 画面のキャプチャー画像を対象として OCR を行う。【画面 OCR】

前者(1)はデータ・エントリー業務の自動化で用いられる OCR であり、スキャナーで読み込まれた文書画像を認識対象とする。後者(2)は PC 画面上に表示されたテキストを認識する OCR であり、PC が生成した画面キャプチャー画像を認識対象とする。

前者の場合、OCR で認識した項目データを RPA が利用するが、OCR の認識結果に誤りがあると自動処理が失敗するため、目視によるデータ確認が行われることが多い（図 2.2-1）。つまり、完全な自動処理ではなく、認識結果には人が責任を持つ必要があるが、それでもキーボードでデータを手入力するよりも作業工数の大幅な削減が期待できる。また、若干の誤りならば許容できる用途の場合は、目視確認を省略することもある。

一方、後者の場合は PC 画面のキャプチャー画像を認識して文字データや座標を連続的に取得し、その座標を用いて自動実行をする。つまり、RPA 実行時に認識するため、人による確認が入る余地が無い（図 2.2-2）。これは OCR が認識誤りを起こすと誤動作に直結することを意味するので、OCR の精度に対する要求レベルは前者に比べてより高くなる。そのため後述するような OCR に影響を与える要因の精査がより重要となる。



図 2.2-1 RPA 向けデータ入力自動化のための OCR 利用



図 2.2-2 UI 操作自動化のための OCR 利用

本報告書では、これまでスキャナー及びカメラで取得した画像を対象とした OCR についてのガイドラインを検討してきたが、近年、比較的新しい OCR の用途として UI 操作の自動化を目的とした画面 OCR を検討対象に加えている。今年度の報告書でもその流れを受け、本節において画面 OCR の精度に影響を与える要因について述べる。(第 1.1 節の俯瞰的考察の図によれば、電子情報がアナログ変換を経ずにデジタルパターン情報に変換され、復元された情報が活用される、図中の①⇒②⇒⑤⇒⑥の処理ルートに相当する)

2.2.1 画面 UI 操作の自動化

RPA における画面 UI 操作は、例えば、画面上のボタンをクリックするなど、操作対象のターゲット (ボタン) と操作内容のコマンド (クリック) を記録したシナリオに基づいて UI 操作を別環境で再生する。そのために多くの RPA ツールはユーザーが画面を操作した内容を記録するコーディング (操作記録) 機能を提供している。(図 2.2-3)

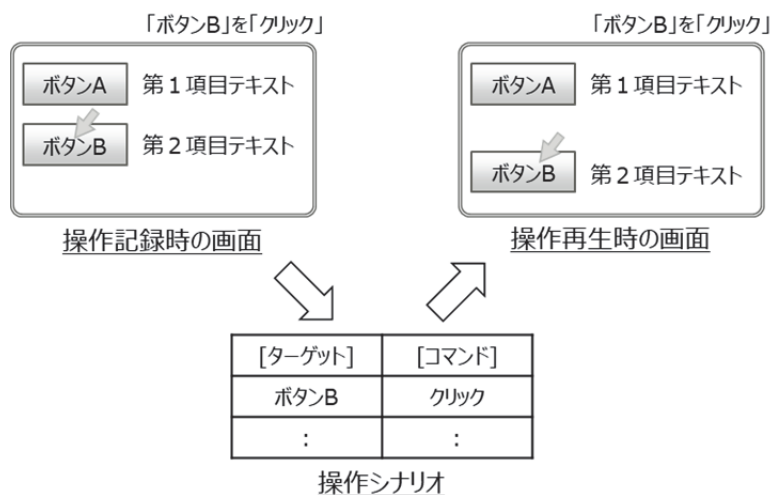


図 2.2-3 RPA における UI 操作の記録と再生

操作対象ターゲットは、例えば、Web ブラウザーの場合には <button id="ボタン B"> のように HTML タグで記録された対象を id や xpath で指定することが多い。また、操作対象の座標値を指定することもある。しかし、図 2.2-4 のように操作対象の id や表示位置が変化する場合には、

操作再生時にターゲットを見つけることができない（操作ターゲットの破損）。また、Web ブラウザ以外のアプリを操作する場合には、そもそも id や xpath によるターゲット指定を行うことはできない。

このような場合の対策として画面キャプチャー画像からターゲットを検索するという手法がしばしば用いられる。例えば、クリックするボタンのアイコン画像を検索してクリック座標を求める方法や、クリック座標の近くにある文字列を検索する方法などがある。その一例を図 2.2-5 に示す。この例では、“第 2 項目テキスト”という文字列を検索し、その左隣にあるボタンをクリックすると記録されている。それを再生時の画面で実行すれば、操作対象ボタンの表示位置や id 等の変化の影響を受けず記録時と同じ UI 操作を実行することができる。ここで文字列を検索するために画面 OCR 技術が使われる。

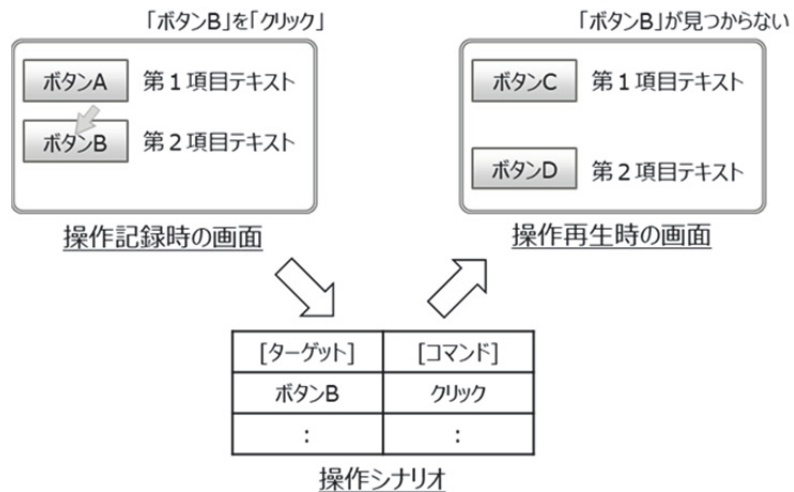


図 2.2-4 RPA における UI 操作の記録と再生（失敗例）

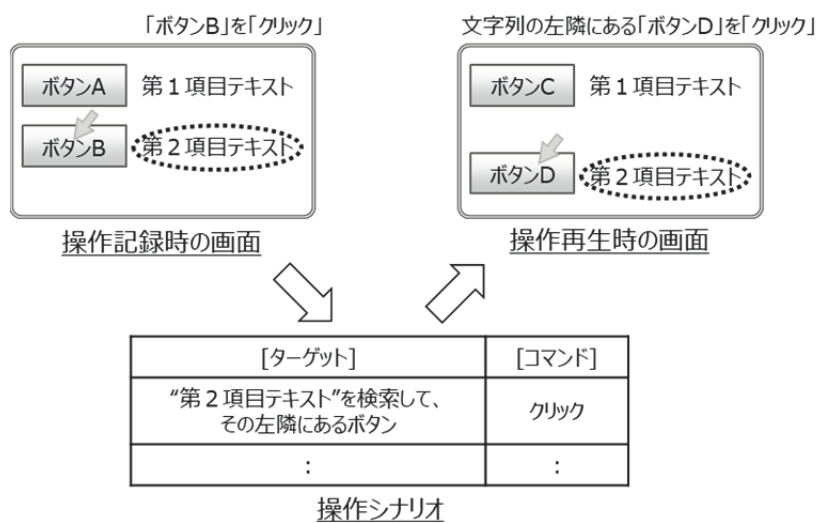


図 2.2-5 RPA における UI 操作の記録と再生（OCR の利用）

参考：

Web アプリケーションを対象とした自動テスト技術では、実行テスト時に行う画面操作を繰り返し自動実行するため RPA と同様の技術が用いられている。操作対象ターゲットを特定する機能は **locator** と呼ばれ、3 種類に分類される（OCR を用いる方法は第三世代に相当）。この研究分野では、3 種類の **locator** の併用や相互変換*によりターゲットの破損を修正する技術（自動修復：Self-Healing）も提案されている。

1. 第一世代 **locator** 操作対象を画面上の座標値で指定する
2. 第二世代 **locator** 操作対象を **Tag** の **id** や **xpath** などの属性値によって指定する
3. 第三世代 **locator** 画面表示画像から画像処理技術によって検索することによって対象を指定する

* “Automated Generation of Visual Web Tests from DOM-based Web Tests”, Maurizio Leotta, Andrea Stocco, Filippo Ricca, Paolo Tonella (SAC 2015, Apr. 2015)

2.2.2 画面 OCR に影響を与える要因

画面キャプチャー画像を対象とした OCR では、図 2.2-6 に示すような撮像過程において画像の質的変換が行われるため、OCR の精度に影響が生ずる。画面 OCR の実行条件を設定するにはこれらの要因を考慮する必要がある。

実世界環境における OCR では文書面（紙など）に当たった照明の反射光を検出して文字画像が生成されるが、画面 OCR では文書面で生成される文字画像は電子的に生成されたものなので、文字画像には変動要因はほとんど無いように思える。しかしながら、図 2.2-6 に示すように、文字画像は文書データからフォント情報に基づいて生成されるため、例えば、ビットマップフォントかアウトラインフォントか、アンチエイリアスの有無、等により文字画像の品質が変化する。また、画面上には文字以外の図形も表示されており、文字を隠したり、文字に近接してオブジェクトが配置されていたりする（例：罫線が文字に接触）と、文字領域を正確に切り出すのが困難になる。更に画面サイズに応じて画像が拡大・縮小されるなど、拡大率や解像度による画像品質の変化も生じ得る。画面 OCR の精度を維持（保証）するためにはこのような変動要因を考慮する必要がある（表 2.2-1）。

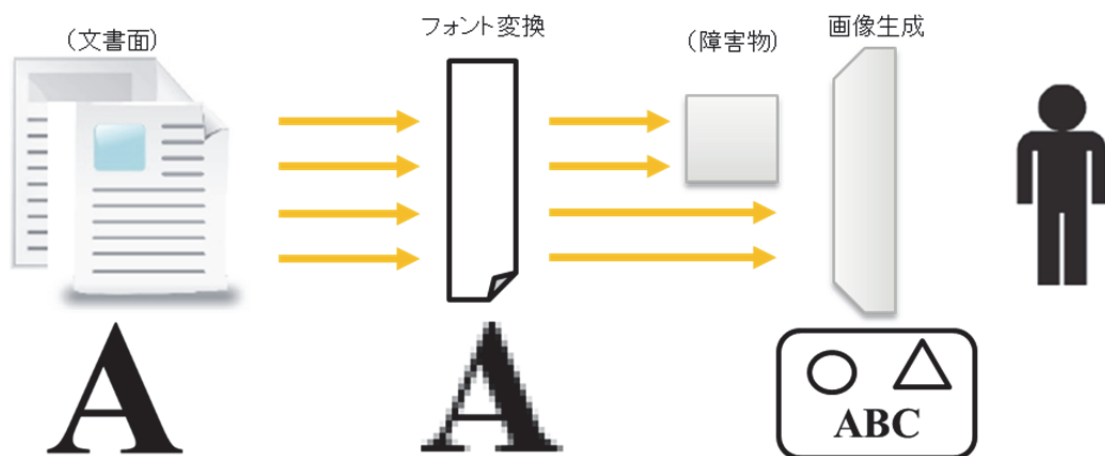


図 2.2-6 画面キャプチャー画像の撮像過程に影響を与える要因

表 2.2-1 画面 OCR に影響を与える要因

分類	要 因
フォント 変換	<ul style="list-style-type: none"> ・フォントの種類 <ul style="list-style-type: none"> -アウトラインフォント、ビットマップフォント、画像 ・文字画像生成 <ul style="list-style-type: none"> -変換無し、アンチエイリアス（スーパーサンプリング、マルチサンプリング、等）
障害物	<ul style="list-style-type: none"> ・文字を隠す障害物 <ul style="list-style-type: none"> -文字に重なるオブジェクト、文字間の重なり ・文字抽出の妨げになる障害物 <ul style="list-style-type: none"> -文字以外の近接オブジェクト、背景模様
画像生成	<ul style="list-style-type: none"> ・表示品質 <ul style="list-style-type: none"> -表示拡大率（ディスプレイ、ブラウザー）、画面解像度

2.2.3 RPA における OCR 認識誤りの影響

OCR の認識結果を RPA 実行のための入力とした場合、認識誤りが発生すると動作にエラーが生ずるため、認識誤りを避ける対策が必要となる。そのため認識結果の目視確認を行い、精度保証の責任を人間に任せるといった手段が用いられることが多い（図 2.2-1）が、RPA 実行時に OCR を行う UI 操作自動化においては（図 2.2-5）、OCR を含むシステム自身が精度保証を求められるため、①十分に実用になる程度の認識精度を保証する、②仮に認識誤りが発生してもユーザーの用途にとって致命的な影響を与えないことを保証する、のいずれかが求められる。

以上のような、「認識誤りによるユーザーへの影響の度合いを把握し、適切な品質保証を行う」ことは、パターン認識技術（OCR を含む）を実用化する上で不可欠であり、近年では AI 品質保証の問題として議論が活発化している。詳しくは 2.4 節にて述べる。

2.3 実世界環境における認識機器（技術）の現状と今後の展望

2.1 節で説明しているように、当委員会では OCR の認識精度に影響を及ぼす「実世界環境における外乱要因」について調査検討を継続しているが、OCR は長年の商品化の歴史において、徐々に実世界環境への対応が進んできている。

本節では、2.3.1 で、OCR 関連技術に関係が深い、DL の進化や研究成果に関するトピックスとして、Transformer モデル、ChatGPT、画像生成について現時点での状況を紹介した。2.3.2 では、最近話題になっている AI-OCR について、「従来 OCR と AI-OCR との比較」をした。また、2.3.3 では「AI-OCR における現状の課題」、2.3.4 では「今後の展望」について記述した。

2.3.1 実世界環境における認識機器（技術）のトピックス

2006 年の ILSVRC（Imagenet Large Scale Visual Recognition Challenge）において、DL を利用した手法が高い認識精度を達成したという発表があり、その後も DL は日々進化を続け、色々な分野で認識精度の向上に貢献してきた。時系列データ（音声、テキスト）の処理においても、

RNN (Recurrent Neural Network) や LSTM (Long Short Term Memory) が提案され、2017年には Google から Transformer モデルが発表された。2018年には BERT (Bidirectional Encoder Representations from Transformers) や GPT (Generative Pre-trained Transformer) などが発表されるなど、Transformer ベースの事前学習モデルの利用は増々進んでおり、応用分野も広がっている。

Transformer は層を深くすればするほど、言語モデルでは学習データから抽象的な内容を獲得できることが特徴で、モデルを巨大化することで文章の意味を、より正確に捉えられるようになる。Transformer は、言語や音声処理から Vision and Language、そして、コンピュータービジョンへの適用 (ViT: Vision Transformer) へと進んでいる。CNN (Convolutional Neural Network) ベースのモデルでは物体のテクスチャを頼りに画像を認識するのに対し、ViT は人間のように物体の形状を頼りに認識していることも確かめられ、認識結果の不自然性が減少すると言われていた [1]。最近では、ViT を文字認識に利用する発表も出始め、今後、色々な分野への応用が期待される (3章)。

上記のように DL の進化 (Transformer、など) や学習データの作成手法の進化 (3章) は、認識機器 (技術) の精度を急速に向上させており、スマホ等の音声認識や音声応答、対話システム、翻訳システム、経済や法律、医学などの専門的 AI システムなどにおいて、実世界環境で利用できる精度になりつつある。

最近、Transformer を利用して学習した「生成 AI」が注目されるようになり、テキストを生成する対話 AI: 「ChatGPT」などや、画像を生成する AI: 「DALL・E」などが公開されている。ChatGPT は OpenAI によってトレーニングされた大型の言語モデルで、人工知能により生成されたテキストを返すことができ、会話や質問応答、文書生成など、多様なタスクに使用することができる (ChatGPT 自身の説明)。

ChatGPT は、2022 年の 11 月に公開されて以来、5 日間で 100 万ユーザーを獲得し (前例のない速さ)、色々な分野で性能の評価や使用方法に関する検討などが進められている。自然な対話ができ、翻訳や簡単なプログラム作成やデバッグ、などもできることも示されている。また、ChatGPT は検索技術を一変させるとも言われ、Google などのビジネスモデルへの影響も指摘されているが、Google も「Bard」という「ChatGPT」に似た自社独自の AI チャットボットの評価版を 2 月に発表するなど、開発競争が激化している。

画像を生成できる AI 画像生成システム (DALL・E2、ERNIE-ViLG、Midjourney、Stable Diffusion、など) においても評価版が公開されている。これらは、文章を与えてアニメ調の画像などを生成することなどができる。

生成 AI 技術は、大量の学習データを、Transformer モデルでコンピューターパワーを利用して学習した技術であるが、進化は急速に進んでいる。また、生成 AI 技術は今後、色々なサービス

に組み込まれることが予想され、認識機器（技術）の精度向上や「実世界環境での利用」を可能にしていくと期待される。急速に進化を続けている DL 関連技術については、継続して調査を進めていきたい。

2.3.2 従来 OCR と AI-OCR との比較

本節では、AI-OCR が登場した経緯や、従来 OCR に比較して優れている点、現状の課題などについて紹介する。

1) 従来 OCR（帳票 OCR、文書 OCR）

OCR は、手書き・活字の帳票を読み取る「帳票 OCR」と、活字の一般文書などを読み取る「文書 OCR」が商品化されている[2][3]。

「帳票 OCR」は、読み取りに適した「定型帳票」を作成し、作成した帳票に対応する読み取り方法を定義（帳票定義）し、誤読を抑える工夫をしながら運用する、データ入力ツールとして長年利用されてきた。最近では、帳票 OCR の多くが簡単な帳票解析機能を組み込み、既存帳票も読み取ることができるように進化している。

一方、「文書 OCR」は、新聞、雑誌、書籍、名刺、などを読み取ることが出来る OCR で、文書のレイアウト構造を解析し、テキスト領域における色々な書体の活字文字列を高精度に読み取るツールとして広く利用されている。

「帳票 OCR」、「文書 OCR」共に、読み取り環境や対象帳票・文書に対する実世界環境への対応要望は多く、研究開発は継続的に進められてきた。2017 年頃から、RPA という PC の操作プロセスを自動化するソフトウェアの出現により、非定型帳票の認識や、フリーエリアに書かれた自由手書き文字列の高精度認識、複雑なレイアウトの文書の認識、などを高精度に行いたいという要望が多くなり、対応できる OCR の開発が課題になってきた。

2) AI-OCR

2016 年頃から、国内においても文字・文書認識技術に DL を利用する研究が盛んになり、従来の文字認識精度を上回る成果が出始めた。その後、3 章で報告してきたように、「非定型帳票内の文字列」や「フリーピッチの手書き文字列」の認識などにおいても、DL を利用した方式が従来方式に比較して優れた結果が報告されるようになり、成果は 2017 年ころに「AI-OCR（AI という場合もある）」という名称でリリースされ始めた。最近では、既存の OCR を DL を利用して精度向上した従来メーカーの AI-OCR や、ベンチャー企業の AI-OCR が市場に参入し、OCR 市場は拡大している。AI-OCR が出現した背景には、ユーザー要望（RPA での利用、他）に対して、DL 技術を利用した認識技術の進化と、計算機環境の進化が併行して進んできた結果とも考えられる。

3) AI-OCR における、「帳票 OCR」と「文書 OCR」機能の実現手法例

AI-OCR は、RPA 利用者の要望に対応する OCR として、従来のソフトウェアタイプ「帳票 OCR」と「文書 OCR」の機能を持ち、DL を利用した精度向上（下記、i）、ii）、など）が行われてリリースされてきた。特に、従来 OCR に比較して「非定型帳票内の文字列」や「フリーピッチ手書き文字列」の高認識精度をアピールしていることが特徴と言える。

i) 非定型帳票内の文字列の認識精度を向上する手法例

2020 年度の本委員会報告書[2]に、DL を利用した領収書の認識方式例を示したが、その例では認識画像領域を、DL（CNN など）を利用して抽出し、抽出した画像領域を DL で文字認識して結果を出力するという方法などがとられている。最近、帳票識別機能を持った AI-OCR も商品化されているが、帳票の一部や全体を、DL（CNN 等）を利用して識別していると予想される。

ii) フリーピッチ手書き文字列の認識精度を向上する手法例

2020 年度の本委員会報告書[2]では、DL を利用して、文字列の個々の文字の中心位置を検出し、その文字矩形を見つけて認識する手法を紹介した。また、2021 年度の報告書[3]では、文字列画像を一定間隔で切断し、切り出した文字片画像を CNN で認識し、その結果を RNN や LSTM に入力して文字片画像単位に出力し、認識した文字列を、CTC decoder を利用して最終的な文字列に変換する、End to End 方式を紹介した。両方式とも、DL を利用して実現できた方式と言える。

2.3.3 AI-OCR における現状の課題

AI-OCR は、従来 OCR に比較して認識精度が向上したと言われているが、従来からの課題がすべて解決したわけではなく、今後の改良に対する期待も大きい。当面の課題は、2.3.2 の 3) - i)、ii) で紹介した項目のさらなる精度向上、及び確認修正の効率化などが考えられる。

1) 非定型帳票内の文字列認識、複雑なレイアウトの文書認識

色々な帳票・文書に対して、従来 OCR の課題である、帳票解析・理解、文書の論理構造解析・理解などは、DL を利用しても短期間で解決することは難しい。世の中に存在する、帳票の「表形状」や「文書構造」は多種多様で、現状では案件ごとに対応（学習など）しながら導入している場合も多く、今後の課題の 1 つと考えられている。

2) フリーピッチの手書き文字列の認識

住所や氏名などの定型文字列に関しては、長年蓄積してきた大量データを、DL を利用して学習することにより、従来 OCR を上回る高精度な AI-OCR がリリースされている。しかし、自由に書かれた手書き文字列に関しては、大量の学習データの収集が難しい場合も多く、短期間での精度向上は難しく、今後の課題の 1 つと考えられる。

3) 読取り結果の確認修正の効率化（誤読をゼロにする仕組み）

帳票読取りに利用されてきた従来の「帳票 OCR」は、「人手による入力」と「OCR を利用した場合の入力」を比較し、入力コストが下がることを確認した上で導入されてきた。商品カタログで認識率 99.0%と表示されている OCR においても、1%は誤読していることになるが、どの部分を誤読しているかまでは正確にわからないことが多い。そこで、すべての認識結果を確認しながら誤読部分を修正する必要があり、入力作業の効率化に繋がらないこともある。

従来の「帳票 OCR」では、認識結果をすべて確認しないで済む機能が用意されている。その例として、OCR での認識結果に自信が無い、あるいは認識不能な場合は、認識結果を文字や文字列（単語など）単位でリジェクトとし、確認修正はリジェクトした部分を中心に行うという機能である。また、帳票内にリジェクト部分（文字、文字列）がある場合には、その帳票自体をリジェクトとし、リジェクトした帳票だけを確認修正するという方式もある。これは、リジェクトした帳票以外の帳票の誤読を限りなくゼロにするという考え方で、一般に通過率と呼ばれている。通過率はリジェクトを含む帳票を除外し、通過した帳票が全体帳票の何%であるかを計測するもので、確認修正に必要な時間を推定する指標の1つとして用いられる。例えば、OCR の通過率が 90%であれば、1,000 枚の帳票を読み取った場合、確認修正する帳票は 100 枚（10%）ということになる。

OCR の認識結果全体を確認修正するのではなく、リジェクトした部分や帳票のみを確認修正することで、入力コストを削減する機能は、AI-OCR においても重要な検討課題と考えられる。下記（引用 1）、（引用 2）に、認識率、誤読率の定義を、[4]から引用して記述した。

-----（以下、引用部分）

（引用 1）認識率の定義 [4]

OCR の精度を評価する方法として、実際に画像に記載されている文字と、OCR 処理された結果を比較して行う認識率の評価がある。OCR で誤認識してしまう要素には、下記の種類がある。

- ・認識して欲しい文字とは違う文字で認識されている（誤った文字で認識されている）
- ・認識して欲しい文字が出力されていない（認識もれがある）
- ・ unnecessary 文字が含まれている（不要な文字が挿入されている）

認識率の計算方法は主に下記二種類が利用される。

- ・ 1 文字ずつ比較して計算する（文書全体の認識率）

（正しく認識された文字数／画像上の文字数）×100＝認識率(%)

- ・ フィールド単位、ページ単位で比較して計算する（完全一致の認識率）

（文字がすべて正しく認識されたフィールドもしくはページ数／すべてのフィールド数も

しくはページ数) ×100=認識率(%). この方式を通過率と呼ぶ場合もある。

記入されている文字を他の文字として認識してしまったものを誤読、あるいは誤認識という。

(引用 2) 誤読率の定義 [4]

誤読した文字数 / 全体の文字数 × 100 = 誤読率 (%)

誤読が発生した場合、すべての誤読文字の訂正処理が必要なため訂正するための負荷がかかる。そのため、業務によっては誤読よりリジェクトが好まれ、認識率より誤読率が重要視される場合がある。

----- (ここまで)

今後、上記のような課題は徐々に解決していくと思われるが、効率よく改良していくには、長年研究開発を進めてきた OCR 関連技術 (帳票解析・理解、文書の論理構造理解、など) や、OCR の運用ノウハウなどは、参考になると考えられる [4][5]。

2.3.4 今後の展望

実世界環境における認識機器 (技術) のトピックス、AI-OCR が普及してきた経緯、AI-OCR の特徴、現状の課題、今後への期待について述べた。特に、「帳票 OCR」は長年の歴史の中で、人手による入力 (人力入力) とコストパフォーマンスを比較されながら進化してきた。また、認識結果の確認修正に関しては、OCR を運用するときの手法やノウハウとして、各社が地道な対応を進めてきた。今後、OCR をさらに普及させていくためには、これらの手法やノウハウを参考にしながら、DL 等の新技術を利用した機能・性能向上を併行して進めていくことが必要と考えられる。

色々な帳票や文書に関して、OCR の認識精度を 100% にすることは短期間では難しいと予想されるので、当面はシステム全体の運用を考えた検討も必要になると思われる。特に、確認修正の効率化は OCR を導入するための重要な課題の 1 つと考えられてきた。参考文献 [4][5] は、長年の OCR の歴史の中で開発された技術や機能を「OCR カタログ用語集」や「OCR ガイドブック」としてまとめたもので、参考になると思う。

RPA の普及や電子帳票保存法の施行等で、OCR の利用範囲は拡大している。この機会を前向きにとらえ、多くの企業やユーザーに導入してもらえるデータ入力ツールとして、AI-OCR が増々普及していくことを期待したい。

【参考文献 (URL は 2023 年 3 月 17 日時点)】

- [1] Shikhar Tuli, et al: 「Are Convolutional Neural Networks or Transformers more like human vision?」 CogSci, 2021
- [2] 認識形入力方式に関する調査研究報告書 2021 年 3 月 JEITA (情産-21-技標-1)

- [3] 認識形入力方式に関する調査研究報告書 2022年3月 JEITA（情産-22-技標-1）
- [4] OCRカタログ用語集 第3版 2022年1月 JEITA OCR 専門委員会（IS-09-情端-7）
https://home.jeita.or.jp/upload_file/20130502145640_LBIW8P10de.pdf
- [5] OCRガイドブック ～OCRを上手にご活用いただくために～ 2010年5月 JEITA OCR 専門委員会 <https://home.jeita.or.jp/is/ebook/>

2.4 OCRの品質保証

2.2節で述べたように、OCRの認識結果を自動処理に用いる場合は認識誤りによる誤動作リスクが問題となる。OCRをはじめとするパターン認識技術を実用化するためには、誤認識によるダメージを許容範囲内に収める必要があるが、どの程度のリスクなら許容可能かも含めたリスク管理の考え方はパターン認識の分野では十分に浸透していない。そのため、ソフトウェア開発におけるリスク管理の基準をそのまま適用し、例えば、「0.01%でも誤動作の可能性があるのならテスト不合格」、「誤認識はソフトウェアの誤動作であり、瑕疵とみなす」のような極端な保証要求が設定される場合があるなど、パターン認識技術の実用化を阻む要因となっている。

近年では深層学習の技術的な発展に伴うAIシステムの普及により、信頼できるAIの実現が広く求められている。識別器としてのAIはOCRやパターン認識を包含する概念であり、OCR実用化のための品質保証問題は、そのまま「AI品質」の問題として議論・検討することができる。AI品質保証の問題は、更に広くAIを適正に利活用するためのガイドライン策定の動きとも絡んでおり、国内外の公的機関が様々な指針を提示している。それは単にAIシステムの動作の安定性・確実性といったレベルの問題から、AIを用いることの社会的な影響など、より広い視点での議論や考察を含む。下記に主な国内のガイドライン策定プロジェクトと、AI実用化に向けた研究開発プロジェクトを挙げる。その他にも安心して使えるAIの普及を目指した動きは多く、本稿ではそれぞれの動きの詳細まで記すことはできないので、興味がある方は参考文献を参照して欲しい。（IT各社はそれぞれAI利用における信頼性や責任について企業としてのスタンスを宣言していることが多い。参考文献[13]以降にてその一部を紹介する）

■ 機械学習品質マネジメントガイドライン（通称：産総研AI品質ガイドライン）[1]

<https://www.digiarc.aist.go.jp/publication/aiqm/>

NEDOの委託に基づいて産総研が主導して作成した、機械学習の品質マネジメントに関するガイドラインである。企業・大学などの外部有識者とともにまとめたもので、AIを用いた製品やサービスの品質を安全、安心に管理するための指針を提供する。

ガイドライン本編は2022年8月に第3版（第3.1.0版）がリリースされ、2023年1月に改版された第3.2.1版が最新である。英語版も2023年1月に第3版が公開された。また、

ガイドラインの具体的な使用法の例を記した「機械学習品質マネジメントリファレンスガイド」も別途提供されており、その中に郵便番号 OCR の事例も記述されている。

■ AI プロダクト品質保証ガイドライン（通称：QA4AI ガイドライン） [2]

<http://www.qa4ai.jp/download/>

QA4AI とは、2018 年 4 月に設立した産学コンソーシアムである。Web ページの説明によれば、“AI プロダクトの品質保証に関する調査・体系化、適用支援・応用、研究開発を推進するとともに、AI プロダクトの品質に対する適切な理解を社会に啓発する活動を行うコンソーシアム”であり、“AI 技術の活用・進化のさらなる促進と、AI プロダクトと社会との安心できる共生の実現を”目指すとされている。

QA4AI が作成・公開している AI 品質保証のためのガイドラインが「AI プロダクト品質保証ガイドライン」であり、2019 年 5 月に初版がリリースされ、現時点で 2022 年 7 月版が最新である。

■ Engineerable AI プロジェクト（通称：eAI プロジェクト） [3]

<https://engineerable.ai/>

<https://www.jst.go.jp/mirai/jp/program/super-smart/JPMJMI20B8.html>

eAI プロジェクトは、国立情報学研究所の石川冬樹准教授を代表として、AI の安全性、信頼性を確立する“Engineerable AI(eAI)”という汎用的基盤技術の開発を目指した、JST 未来創造事業に採択されたプロジェクトである。2020 年の探索研究を経て、2021 年 1 月から本格プロジェクトが開始している。主に自動運転 AI と医療診断 AI をターゲットとして、信頼できる AI システムの構築技術の検討が行われている。

【参考文献】

- [1] 機械学習品質マネジメントガイドライン（通称：産総研 AI 品質ガイドライン）

<https://www.digiarc.aist.go.jp/publication/aiqm/>

- [2] AI プロダクト品質保証ガイドライン（通称：QA4AI ガイドライン）

<http://www.qa4ai.jp/download/>

- [3] Engineerable AI プロジェクト（通称：eAI プロジェクト） <https://engineerable.ai/>

- [4] 我が国の AI ガバナンスの在り方 ver1.1（経産省 2021 年 7 月 30 日）

https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/2021070901_report.html

- [5] AI ネットワーク社会推進会議（総務省）

<https://www.soumu.go.jp/iicp/research/results/ai-network.html>

- [6] AI 応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立 (JST CRDS 2018 年 12 月) <https://www.jst.go.jp/crds/report/CRDS-FY2018-SP-03.html>
- [7] 人間中心の AI 社会原則 (内閣府) <https://www8.cao.go.jp/cstp/aigensoku.pdf>
- [8] Trust and Artificial Intelligence (NIST)
https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931087
- [9] ISO/IEC JTC 1/SC 42: Artificial Intelligence
<https://www.iso.org/committee/6794475.html>
- [10] Explainable Artificial Intelligence (XAI)
<https://www.darpa.mil/program/explainable-artificial-intelligence>
- [11] Ethics guidelines for trustworthy AI
<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- [12] AI ガイドライン比較表(総務省) https://www.soumu.go.jp/main_content/000630126.pdf
- [13] 責任ある信頼された AI (マイクロソフト)
<https://learn.microsoft.com/ja-jp/azure/cloud-adoption-framework/innovate/best-practices/trusted-ai>
- [14] Google と AI: 私たちの基本理念 (Google)
<https://japan.googleblog.com/2018/06/ai-principles.html>
- [15] 富士通グループ AI コミットメント <https://pr.fujitsu.com/jp/news/2019/03/13-1a.pdf>
- [16] 東芝グループ AI ガバナンスステートメント
<https://www.global.toshiba/jp/technology/corporate/ai-statement.html>
- [17] NEC グループ AI と人権に関するポリシー
<https://jpn.nec.com/press/201904/images/0201-01-01.pdf>
- [18] リコーグループ AI 活用基本方針
https://jp.ricoh.com/-/Media/Ricoh/Sites/jp_ricoh/security/pdf/basic_policy_for%20AI_technology_utilization_japanese.pdf

3. 認識技術の動向

3.1 文字認識技術の現状と今後の展望

本節では、主に企業におけるビジネス適用の観点から、文字認識／DL 技術の動向について述べる。

3.1.1 文字認識技術概説

OCR に代表される文字・文書認識は、パターン認識研究の最初期から取り組まれている研究課題である。パターン認識分野における様々な基礎的技術が文字認識研究を通して開発され、他の領域へと応用されてきた。文字認識分野では、ETL9 や MNIST データセットに代表される認識手法を評価するための標準的データセットが早期に整備されており、長年にわたりパターン認識や機械学習研究で提案される種々の認識手法のベンチマーキングに活用されている。

図 3.1-1 に一般的な文字認識処理の流れを示す。文字認識は書面全体、又は書面に書かれた文字を各種マークアップファイルや文字コードで表現されるデジタルデータに変換する処理であり、主な構成要素は以下に示す 6 要素である。

1. 画像入力：文書・書面など、文字が記された現実媒体を、スキャナやカメラなどを通じデジタル画像として計算機に取り込む処理
2. 前処理：文書画像のノイズ除去、コントラスト強調、二値化などのデジタル画像に含まれる文字を見やすくし、文書のレイアウトを解析する、モデル定義に従うなどして行の切り出しを行う処理
3. 文字切出：文字列切出によって切り出された文字行から 1 文字ずつの切り出しを行う処理
4. 特徴抽出：文字の見かけを表現する特徴量を抽出し、特徴ベクトルを構成する処理。抽出された特徴ベクトルをより識別に有用な特徴ベクトルに変換する（例えば、計算量を削減する目的で特徴ベクトルの次元数を削減する）処理を加えることもある
5. 分類：入力された特徴ベクトルに対し、あらかじめ学習済みの分類機を用いて文字ラベルを割り当てる処理
6. 後処理：辞書データや言語モデルとのマッチングなどにより認識結果を補正して認識精度を向上させる処理

最近では DL 技術の進展に伴い、多くの場合これら処理の一部、あるいはすべてが DL 化されている。DL の代表的な利用例としては特徴抽出・分類を置き換える場合（図 3.1-1 中②組み合わせ）、文字切出・特徴抽出・分類を置き換える場合（図 3.1-1 中③End-to-End）の 2 例がある。

特徴抽出・分類を置き換える場合には CNN (Convolutional Neural Network) を、文字切出・特徴抽出・分類を置き換える場合には RNN (Recurrent Neural Network)、特にその拡張である LSTM (Long Short Term Memory) をベースとする手法が用いられることが多い。これらの手法は複数の処理を同時に学習することで、後段の処理 (分類) にとって都合のよい前段の処理 (文字切出・特徴抽出) が学習データから自動的に獲得されることが大きな特徴である。このため十分なデータを学習することができれば、人手による詳細な設計を必要とするこれまでの処理を用いた場合よりも容易に、高い認識精度を得ることができる。複数の処理を一括で行う手法としては以前から HMM (Hidden Markov Model) などが知られていたが、手書き英字の認識など一部の分野でのみ用いられていた。しかし近年、DL 技術が急速に進展して認識精度が大幅に向上したこと、計算機リソースを豊富に使えるクラウド環境上でサービスとして提供する形態への変化が進んだことから、文字認識製品で基本技術として採用されるのは End-to-End の DL が中心となっている。

最近では明示的に文字認識を行わず画像から直接 (多くは各文に属性が付与された) テキストに変換する、あるいは画像 (群) に関する質問への応答として必要な情報を得るといった、より End-to-End な方法で情報抽出を行う研究もなされている。今後もしばらくは技術的な革新が続くと予想され、引き続き注目を要する。

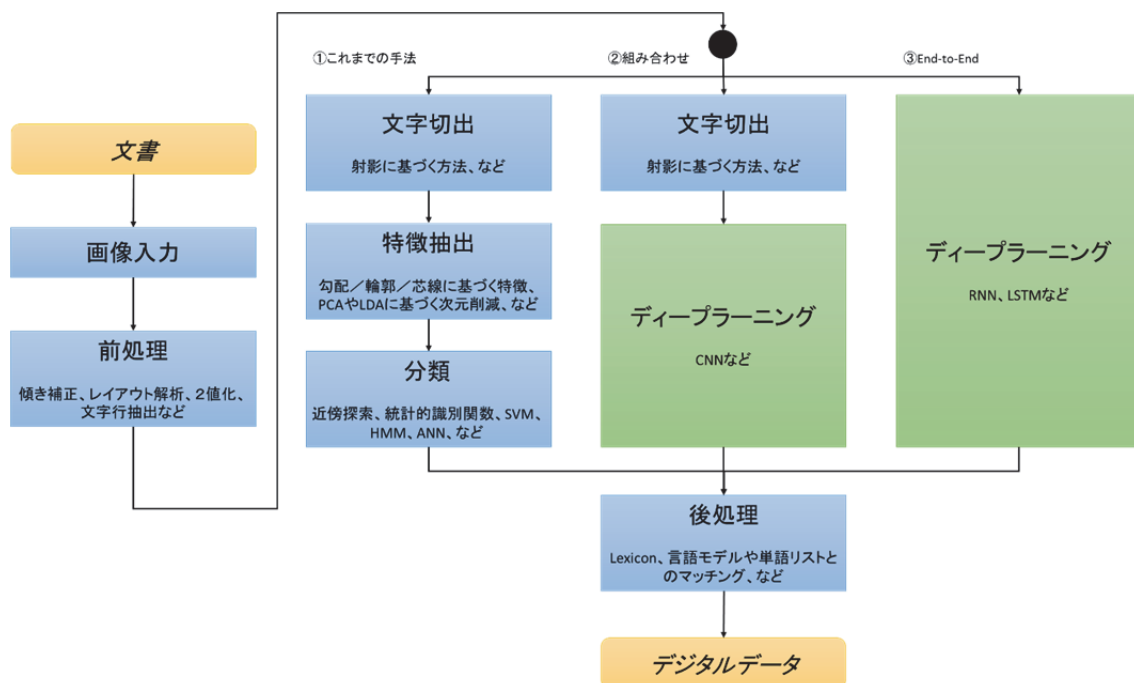


図 3.1-1 文字認識技術の概略

3.1.2 DL/人工知能技術を用いた OCR

DL に代表される AI 技術の進展により一文字ごとの文字記入枠を設けないフリーエリアの認識

率が大きく向上して実用水準に近付いたことで、多くの会社が手書きフリーエリアの認識を含む帳票認識を主な対象としたサービスを一般に提供している。そのようなサービスの代表的な例として、AI-Inside の DX-Suite や Cogent Labs の Tegaki が知られている。これらの OCR はこれまで長年用いられてきた OCR と差別化するために AI-OCR と呼ばれており、DL 技術により高精度を達成したことが特徴である。例えば、Tegaki では活字文字認識製品と同程度以上と考えられる 99.22% の認識精度を達成したと謳っている。

文字認識は機密性の高いデータを読取対象とすることが多く、このため昨今はセキュアなオンプレ環境での提供が拡がりを見せている。DX-Suite では AI Inside Cube によるオンプレミス版がすでに提供されていたが、Cogent Labs も同社の次世代 OCR である SmartRead をオンプレミスに対応させた。本製品では特に高い安全性が求められる金融・公共・BPO 等での業務効率化に貢献することが謳われており、これらの業界を中心に今後ともオンプレミス提供は一つの重要な要素になると思われる。オンプレミスでは株式会社 EduLab の DEEP READ の CPU 版のように高価な GPU を必要としないことを謳うバージョンも出始めており、今後既存のサーバを利用するオンプレミス提供などで利用が広がっていく可能性がある。

利用形態の拡大も進んでいる。代表的な例として、スマホを入力デバイスとしたカメラ型 OCR 「EdgeAI-OCR」を活用した AI 検品ソリューションである。本ソリューションでは、スマホなど小型端末で手書き日本語、バーコード、賞味期限、型番、製造番号などの商品情報を読み取ることで入出荷などの業務を効率化する。このような、現在ではほぼすべてのビジネスパーソンが携帯し、スペックも高まってきているスマホを様々な用途に活用するサービスは今後も進展していくことが予想される。

全体的に AI-OCR は順調に発展しているが、認識精度の点で大きな改善は見られず、利用しやすいサービスへの展開が進んでいる傾向がみられる。既にあらゆる利用シーンにおいて認識精度が十分であるわけではないが、こと手書き文字を中心とした定型帳票認識においては、認識精度改善のフェーズは一段落しているように思われる。

3.1.3 多様なデザインの帳票を対象とした OCR

昨今では働き方改革のための業務効率化ツールとして RPA に注目が集まっている。

RPA から頻繁に連携される技術の一つが OCR である。企業内には他社から紙で送られてくる（請求書など）、証跡として印鑑が必要（申込書など）、といった理由から紙に依存した業務が多く残っている。RPA+OCR は、インフラの維持整備コストの観点からすぐに電子システムへ完全移行することはできないが、こういった紙を用いた業務の負荷は少しでも軽減したいというニーズに応えるものである。

こういった用途に答えるために、請求書に代表される、発行元ごとに書式は異なるが内容はほ

ば同一の帳票、いわゆる非定型帳票を読取対象とした OCR が多く提供されている。これらのサービスでは、項目に付与された見出しや項目自体の書式、及びそれらの位置関係といった情報を手掛かりにして必要な項目を自動的に抽出する。このため事前の読取定義、少なくとも読取領域の位置と種別の指定を行う必要はなく、ユーザーにとってはそれらを要する定型帳票認識よりも簡単に導入・利用することができるメリットがある。これらの OCR もまた AI-OCR と呼ばれる。このような帳票の代表例が請求書である。請求書を扱う OCR は 2023 年 10 月 1 日から開始されるインボイス制度対応を進めており、代表的な例である Infordio の smartOCR では電子帳簿保存法・インボイス制度に特化したサービスの API 提供を開始している。必須要件となる消費税率別記載に対応するなどしており、API であることで顧客システムに容易に組み込み可能であるとしている。並行して読取対象の拡大や顧客業務を踏まえた機能強化も継続的に進展しており、例えば、Cogent Labs. の SmartRead では給与支払報告書対応、FAX 帳票の読み取り強化などを精力的に進めている。

また、これまで進んできた RPA ツールなどとの連携に加え、データ基盤と AI-OCR の連携も進みつつある。例として DATAFLUCT 社のデータ基盤「AirLake」が、東芝デジタルソリューションズの「AI OCR 文字認識サービス」及び「RECAIUS ナレッジプラットフォーム」との連携を発表している。AirLake に蓄積したデータを AI OCR によりデジタル化、さらにナレッジプラットフォームで各事例の特徴を抽出することで蓄積したデータをワンストップで活用することができるとしている。

今後は OCR によるデータ化のみならず、その後の業務との連携、あるいはデータ活用を含めたトータルなソリューションとしての提供がますます進展していくと思われる。

3.1.4 AI-OCR の拡がり と 定着

AI-OCR の普及・発展に伴い、様々なシーンに OCR が登場するようになってきている。これまでも Kaggle コンペティションなどでは OCR の課題が用いられてきたが、今年に入って SIGNATE で凸版印刷主催のくずし字認識コンペティションが開催された。凸版印刷は同社のくずし字 OCR 技術に最先端の AI 技術を取り入れて最高品質のくずし字 OCR 技術実現を目指すとしており、オープンな環境を活用した企業の技術革新に向けた取り組みとして注目される。

OCR が AI 教育の題材となる例も増えてきている。代表的な例として AIJobColle が開催した OCR をテーマとしたハンズオンがある。古くより OCR は画像認識の代表的な例として知られており、実ビジネス化が進んでいる分野でもあるため、AI 教育の題材としては今後も多く活用されていくと予想される。

一方で、OCR の課題に関しても発信が増えている。株式会社ハンモックによれば、AI-OCR の導入後の運用について、運用担当者の約 8 割が課題ありと回答している。中でも課題として挙げら

ている上位 3 項目が、「確認作業の手間がなくなる」「文字認識の精度が低い」「様々な帳票に対応できない」という点である。手書きを中心とした文字認識の精度、多様な帳票への対応は AI-OCR 化により大幅に改善された点ではあるが、利用者視点で見ればまだ十分に精度が高いとは言えないことが示されている。また、確認作業の手間がなくなる点は AI 化以前から続く OCR の課題で、OCR に限らず認識は 100%の精度に達することはないため技術強化のみでは解決が困難であると考えられている。疑わしい認識結果を特定する文字リジェクト機能など確認作業を削減する工夫を適用するなど、ソリューション全体として解決を図っていく必要がある。

【参考文献】

- [1] Gated Recurrent Convolution Neural Network for OCR J. Wang, NIPS2017.
- [2] DX Suite <https://inside.ai/dx-suite/>
- [3] Tegaki <https://www.tegaki.ai/>
- [4] セキュアな環境での文書のデジタル化を推進する次世代 AI OCR「SmartRead On-Premises」の提供を開始 https://www.smartread.jp/news/pressrelease/onpremises_20220907
- [5] より幅広いお客様に手軽にご利用いただける CPU 版「DEEP READ」の発売開始 <https://www.edulab-inc.com/news/press/1119/>
- [6] フューチャーアーキテクト、Edge AI ソリューションを提供開始 https://www.future.co.jp/press_room/PDF/c2673f5140e2b1dbfc8899424e489ceec682986.pdf
- [7] 会計システム向けに電子帳簿保存法・インボイス制度に特化した AI-OCR の API を提供開始 <https://prtimes.jp/main/html/rd/p/000000040.000027032.html/>
- [8] コージェントラボ、IDP サービス「SmartRead」が給与支払報告書の読み取りに対応 https://www.smartread.jp/news/pressrelease/payroll_202206
- [9] 次世代 AI OCR「SmartRead」における FAX 文書の読み取り精度を大幅に強化 <https://www.smartread.jp/news/pressrelease/fax-202208>
- [10] DATAFLUCT のデータ基盤「AirLake」が 東芝デジタルソリューションズの「AI OCR 文字認識サービス」、「RECAIUS ナレッジプラットフォーム」とシステム連携 <https://datafluct.com/news/nzzf4hvyiea>
- [11] SIGNATE にて、凸版印刷主催くずし字認識コンペティション開催！ <https://signate.co.jp/news/2022/202203221300.html>
- [12] OCR を実践的に学びたい方へ！エッジテクノロジー（AI ジョブカレ）が 12 月 21 日(水)に特別講義「画像の分類、手書き伝票の画像認識アプリ」を開催。 <https://prtimes.jp/main/html/rd/p/000000086.000027765.html>

[13] 【AI OCR 導入の実態調査】導入後、約 8 割が課題あり。うち 6 割は「確認作業の手間がなくなる」と回答

https://www.hammock.jp/defact/media/media_pe-less_ocr.html

(URL 確認 2023 年 2 月 12 日)

3.2 文字認識・文書理解に関する国内学会の発表動向

文字認識・文書理解に関する研究報告の調査結果を報告する。調査対象は日本国内で開催された情報系の総合大会・学会・研究会であり、調査期間は 2022 年 1 月から 2022 年 12 月までである。

A. 文字認識

[A2 : 識別／学習]

識別／学習に関する発表が 2 件あった[5][18]。

文献[5]は低解像度ナンバープレート数字認識において、モーメント特徴量によるベイズ識別、スパース主成分分析 (SparsePCA) を用いたベイズ識別、CNN による識別の計 3 つ機械学習モデルの出力に対して、Soft Voting を行った。実験ではナンバープレート中の「3、8、9」に対して識別を行い、Soft Voting によって第二候補まで絞り込めることを確認した。

文献[18]は Vision Transformer (ViT) による文字クラス認識と文字スタイル認識を行い、認識タスクごとに文字のどの部分が重要視されているのかを調べた。実験では画像中の貢献箇所を可視化することで、文字クラス認識時は文字の特徴的な構造部分に、文字スタイル認識時は装飾の施される部分に注目していることを確認した。

[A3 : 文字列認識]

文字列認識に関する発表が 5 件あった[3][11][15][17][19]。

文献[3]、文献[17]は同じ著者による一連の研究である。文献[17]は、Transformer を用いた横書きと縦書きを含む文字列認識において、書字方向を表す特別なトークンをデコーダに入力することで、認識精度を向上させる手法を提案した。実験では、書字方向によって別々に学習したモデルや、書字方向の情報を画像のチャンネル方向に付け加える手法と比較して、認識精度が高くなることを確認した。文献[3]では、文献[17]の手法をベースに、Transformer で文字数を推定する施策を追加し、実験において認識精度向上に寄与することを示した。

文献[11]は中国語手書き文字行認識タスクに対し、事前学習として対照学習を行う手法を提案した。対照学習には online triplet loss を使用した。実験において、提案法は事前学習を行わないモデルと比較して文字誤り率が改善されることを確認した。

文献[15]は、学習データの手に入りにくい近代日本語文書の認識率向上のため、ドメインの異なる合成活字データを、ドメイン適用して学習する手法を提案した。ドメイン適用のため、記載内容の同じ近代文書の文字画像と合成活字画像で、self-attention の特徴ベクトルが近くなるような損失項を追加した。実験の結果、認識精度を 2.3%改善できたとしている。

文献[19]では、漫画中のオノマトペ画像約 1 万枚からなる COO データセットを作成し、既存手法を用いて文字列検出、文字列認識の評価を行った。漫画中のオノマトペは、文字列が極端に湾曲している、文字サイズのばらつきが大きい、といった特徴がある。文字列検出手法として ABCNet v2(*1)及び MTS v3(*2)、文字列認識手法として TRBA(*3)を用いた評価を行い、これら既存の手法では高い認識率を得ることは難しいことを確認した。

*1 <https://github.com/aim-uofa/AdelaiDet>

*2 <https://github.com/MhLiao/MaskTextSpotterV3>

*3 <https://github.com/clovaai/deep-text-recognition-benchmark>

[A5 : オンライン文字認識]

オンライン文字認識に関する発表が 1 件あった[9]。

文献[9]は手書きストロークの時系列を入力とする数式認識において、隣り合う記号同士の結合に対して定義された規則に従って認識結果のグラフを生成する T2D-SCFG を用いることで、不自然な構造の数式が認識結果として生成されることを防ぐ手法を提案している。実験の結果、従来手法では括弧が開いたまま閉じられていない形の不自然な数式として認識されていた事例が、提案手法では正しく認識できるようになったことを確認した。

[A6 : その他]

文字認識に関する研究発表のうち、上記のいずれにも分類されないものが 3 件あった[7][20][22]。

文献[7]はレンダリングされた手書き数式画像からスキャンされた手書き数式画像を生成する CCycleGAN を提案した。スキャンすることによって生じるノイズ等を再現した画像を生成することが可能である。実験では、提案法によって生成された画像を使って学習を行った認識器による数式認識率を検証した。結果、提案法の学習に使用する画像を 100 枚まで減らしても、数式認識率が 30%を超えることを確認した。

文献[20]では、手書き数式認識の精度向上のため、既存の手書き数式認識モデル BTTR の出力候補から正解を選ぶための事後処理として、Siamese Network を用いた距離学習と、候補の集合を考慮した距離学習の 2 つを提案している。集合距離学習の方が精度は高く、評価データでの認識率は 40%程度となった。手書き画像を候補画像の線形和で表したときに重みが負になる場合、正しい候補画像を選択できない例を確認した。これらに対応することを今後の課題としている。

文献[22]では、情景画像中の装飾や傾きを加えられた文字の認識精度向上のために、文字認識の結果を修正することを検討した。認識結果の単語に対して埋め込み表現を計算し、他の単語の埋め込み表現とのユークリッド距離を認識時の単語の編集距離に加えることを提案している。英語のデータセットにおいて実験を行い、約 60%の割合で修正結果が改善されることを確認した。

B. 文書理解

[B1：帳票処理・認識]

帳票処理・認識に関する発表が 1 件あった[16]。

文献[16]は帳票文書を対象に、言語モデルを用いて必要な情報を自動抽出する技術の改良について提案している。提案手法は、特徴抽出器 BERT と分類器 LSTM によって構成され、さらに記号列を特殊トークンに変換する改良と、位置特徴を付加的に用いる改良を加えている。実験の結果、改良を加えないベースラインと比べ、すべての改良を適用した提案手法の再現率、適合率ともに最も良いことを確認した。

[B3：文字を含むカメラ撮像画像解析]

文字を含むカメラ撮像画像解析に関する発表が 1 件あった[6]。

文献[6]は書籍のスキャンやカメラ撮影時に生じる湾曲したひずみを補正するために、AIC（赤池情報量基準）を用いて補正のための近似曲線の次数を決定することで過適合を防ぐ手法を提案している。実験の結果、補正後の OCR の認識精度が従来の補正手法に比べて約 35%向上することを確認した。

C. その他

[C0：その他]

以上のどれにも分類されない研究発表が 9 件あった[1][2][4][8][10][12][13][14][21]。

文献[1]はパスワード入力に単語の筆記を組み合わせることで、パスワードによる本人確認を強化するシステムを提案している。空間上に単語の筆記を行い、取得した筆記の特徴を用いて深層学習により個人識別を行う。利用者 1 人とシステム側が任意に用意した $k-1$ 人による k 値分類を行うことで、システム利用者が増加しても学習コストを一定に保つことが可能なシステムとなっている。

文献[2]では、SNS に見られる漢字が偏と旁に分割されて表されるなどの崩れた表記（「神」→「ネ申」など）が多い文書に対して、OCR を用いて自動修正する手法を提案している。具体的には、文書を文字列に分割して各文字列から文字画像を生成し、文字画像を OCR にかけて修正文を出力させる。2 文字以上で構成された崩れた表記文字列の約 28%を正しく認識できることを確

認した。

文献[4]は、古文書のくずし字認識における単語単位の結果提示手法を提案した。従来の単一文字認識結果の網羅的な文字列の組合せを、単語辞書と照合することで単語単位の候補提示を可能にした。実験では、従来の単一文字認識と比較して、候補提示の平均順位は低下したものの、上位 10 位内に正解が含まれる提示率は向上することを確認した。

文献[8]は個人の字体を模倣した画像を生成するために、デジタルフォントによる文書画像に対して、コンテンツ特徴抽出器、字体特徴抽出器、画像再構成器を用いた字体変換を行なう手法を提案している。実験の結果、概ね実際の個人の手書き画像に類似した画像を生成することができた一方で、長い文字列では文字が掠れて読み取りが困難となるケースがあることを確認した。

文献[10]は個人の字体を模倣した画像の生成において、個人の筆跡の特徴を表現するスタイル特徴は一意に定まらないことを実験的に示し、文字種ごとに異なる特徴空間における類似度に基づく協調フィルタリングによる生成手法を提案している。実験の結果、提案手法は概ね著者の筆跡を明瞭に模倣できた一方で、類似した筆跡の著者が学習データに含まれない著者の筆跡は十分に模倣出来ないことが確認された。

文献[12]は、自動運転の安全評価フレームワークを参考に、高い信頼性が必要な用途に向けた OCR 品質保証のフレームワークの考え方を示した。OCR は外乱が認知のみと少ないため、動作保証条件の運行設計領域 (ODD) 定義やシナリオ作成を要件定義手順に組み込めると述べている。今後、実際のシステム開発での有効性と、品質評価へ活用などの検証が必要であるとしている。

文献[13]では、手書き解答の認識と自動採点に向けて、紙の問題集へ書き込む場合と同等の筆記ログを収集することを目的とする。多種多様な問題や回答方式に対応するための記述手法について検討した内容を述べ、今後の多様な問題形式における課題と展望を述べている。

文献[14]は、画像中の文字に関する質問応答タスクである TextVQA において、誤答の要因を分析するために、画像、質問文、OCR テキスト、あるいはそれらの一部を入力として正答・誤答を予測するモデルを構築した。実験の結果、質問文と OCR テキストを入力としたときに正答・誤答の予測精度が高くなった。

文献[21]は、くずし字認識において極端にデータ数の少ない文字種を補うため、デジタルフォントによる文字画像をコンテンツ画像、くずし字の文字画像をスタイル画像として、FTransGAN を用いたスタイル変換によるくずし字生成を提案している。実験の結果、提案手法は文書ごとにバリエーションのあるくずし字画像を生成でき、さらに学習データにフォント画像データを加えることで、より柔軟に文書の特徴を反映した画像を生成できることを確認した。

表 3.2-1 論文リスト

文献番号	掲載媒体	巻号	著者	題目
1	情処論	63 (4)		特徴の再訓練を必要としない変更可能な筆記
2	FIT	E-036	秋山 大五郎・松原 雅文	OCR を利用した崩れた表記の自動修正手法の性能評価
3	FIT	CH-012	折橋 翔太・山崎 善啓・内田 美尋・高島 瑛彦・増村 亮	日本語情景文字認識のための文字数推定を用いた横書き縦書き共有モデリング
4	FIT	CH-013	片山 歩希・松尾 賢一	古文書翻刻初学者に向けた文字認識処理と単語辞書を用いた翻刻候補文字列の提示手法
5	FIT	I-007	大江 凌太郎・藤田 和弘・四宮 康治	機械学習モデルのアンサンブルを用いた低解像度ナンバープレート数字の識別
6	信学論	J105-D No. 7 pp. 480-484	土井猛・辻裕之・木村 聡	製本原稿に対する単眼カメラ撮影画像の湾曲補正手法
7	PRMU	vol. 121, no. 427	Huy Quang Ung, Hung Tuan Nguyen, Cuong Tuan Nguyen, Tsunenori Ishioka, Masaki Nakagawa	Visual Constraints for Generating Multi-domain Offline Handwritten Mathematical Expressions
8	PRMU	vol. 121, no. 427	佐々木航真, 中村和晃, 新田直子, 馬場口登	デジタル文書画像に対する字体変換処理に基づく手書き文書クローンの生成
9	PRMU	vol. 121, no. 427	北嶋完基, クーン トゥアン グエン, 清水郁子, 中川正樹	オンライン手書き数式認識のための文脈自由文法における木構造解析
10	PRMU	vol. 121, no. 427	玉谷幸輝, 中村和晃, 新田直子, 馬場口登	文字種ごとの特徴空間を活用した協調フィルタリングに基づく手書き文字クローン生成
11	PRMU	vol. 121, no. 427	Trung Tan Ngo, Hung Tuan Nguyen, Masaki Nakagawa	Self-supervised Contrastive Learning Using Triplet Loss for Offline Recognition of Handwritten Chinese Text lines
12	PRMU	vol. 121, no. 427	田中宏	実質的 100%に向けて ; 文書認識のための品質保証フレームワーク
13	PRMU	vol. 122, no. 314	朝倉僚, グエン トゥアン フーン, グエン トゥアン クーン, 宮澤 寛, 土田洋一, 山本高廣, 伊藤雅充, 堀江利彦, 清水郁子, 中川正樹	手書き解答の自動採点に向けた学習ドリルのデジタル化と筆記解答の収集
14	PRMU	vol. 122, no. 314	中村亘岐, 内田誠一	TextVQA タスクの正答可能性判定
15	PRMU	vol. 122, no. 314	Le Duc Anh, Kitamoto Asanobu	Training Kindai OCR with parallel textline images and self-attention feature distance-based loss
16	信学総大	D-5-1	馮 思萌	言語モデルを用いた帳票文書からの情報抽出の改良

表 3.2-1 論文リスト (続き)

文献番号	掲載媒体	巻号	著者	題目
17	MIRU	IS1-86	折橋 翔太, 山崎 善啓, 内田 美尋, 高島 瑛彦, 増村 亮	書字方向トークンを用いた横書き縦書き共有型日本語情景文字認識
18	MIRU	IS2-69	大峠 仁輝, 内田 誠一	Vision Transformer によるパッチベース文字認識
19	MIRU	IS2-71	Jeonghun Baek, Yusuke Matsui, Kiyoharu Aizawa	COO: Comic Onomatopoeia Dataset for Recognizing Arbitrary Texts
20	MIRU	IS2-72	ピョン ケイジ, 松尾 信之介, 内田 誠一	集合距離学習による手書き数式認識結果の事後補正
21	MIRU	IS3-53	阿部 楓也, 岩井 翔真, 宮崎 智, 大町 真一郎	少数くずし字データ補強のための画像生成に関する検討
22	MIRU	IS3-80	梶原 颯希, 宮崎 智, 大町 真一郎	編集距離に基づいた距離学習による情景画像文字認識結果の修正に関する検討

3.3 パターン認識研究の最新動向

3.3.1 数式から自動学習する AI

人工知能 (AI) 技術の導入が期待される分野は多様であり、特に画像認識の技術は注目を集めています。中でも、今回第三次 AI ブームの目玉となっている機械学習の方法の一つである深層学習はデータから自動で各タスクを解くための特徴量を学習できることにあり「画像データセットをいかに構築して学習するか？」にもその本質がある。画像認識の世界的コンペ **ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)** に端を発し、コンピュータビジョン及びパターン認識分野で最も用いられている **ImageNet Dataset** は当初から学術・教育目的のみに限定され、さらに昨今では人種や性別により不平等なラベルを出力してしまうなど、AI 倫理問題が叫ばれている。

ここで、筆者らは特に事前学習 (**Pre-training**) の文脈において、実画像の利用こそが実は同問題を引き起こしている根本原因であるとの仮定のもと、数式から生成したデータセットにより画像認識 AI を (事前) 訓練する手法を提案した[1]。

画像認識 AI の事前学習とは下流となるタスクに先立ち、基礎的な視覚能力をモデルに獲得させる操作のことであり、一般物体認識タスクかつ 100 万画像を超える **ImageNet** を用いることが主流であった。例として文字認識のタスク **MNIST** に追加学習させる際には、まずは **ImageNet** でネットワーク全体を学習 (**ImageNet** の出力は 1,000 カテゴリの一般物体)、その後出力層のみ入れ替え (**MNIST** の出力は 10 カテゴリの数字) ネットワークを学習する操作である。予め現実世界の物体を汎用的に「見る」機能を有しておけば、追加学習の際にもタスクを効果的に解くことができることが多くなる。なお、**MNIST** は事前学習を用いずともほぼ認識できてしまうので、

通常は単体で学習が困難な画像データセットの精度底上げに用いられることが多い。

さて、数式から画像認識 AI を学習する枠組みは数式ドリブン教師あり学習 (Formula-Driven Supervised Learning) と呼称されている。数式、特にフラクタル幾何から生成されるデータセットは Fractal DataBase (FractalDB) であり、現時点では最も良好な精度を叩き出すとして広く用いられつつある。ここで、下記に FractalDB の基本構成について説明する。なお、手法の説明は文献[2]より一部を引用する。

FDSL のコンセプトは、自然画像を認識するために有効な特徴表現を、自然画像を用いない事前学習から獲得することである。そこで FDSL では、先述したように画像パターンを表現可能な数式から自動生成したデータセットを用いて自然画像認識に有効な事前学習モデルの獲得を目指す。

通常の教師あり学習では収集した自然画像に対して人手でアノテーションを行ったデータセットを用いて学習を行う。一方で FDSL の学習に用いるデータセットは、画像及びそのカテゴリが数式から自動生成される。これにより CNN を学習するために必要な大規模データセットを、人手を介せずに構築することが可能となる。また、数式により自動生成される画像は自然画像とは異なり、人物等が含まれないためにプライバシーや安全性／公平性の観点から権利関係等がクリアなデータセットを構築可能であると考えられる。このことから、少なくとも事前学習の文脈においては自然画像を使用した際に生じる人種差別的なバイアス等を持つことはない。

特に、FDSL における FractalDB はフラクタル幾何学に焦点をおいて構築している。フラクタルとは自己相似性を持つ形状のことで、画像においては大域的な形状と局所的な形状が類似しており、同形状が再帰的に組み合わさることでフラクタルを擬似的に表現している。

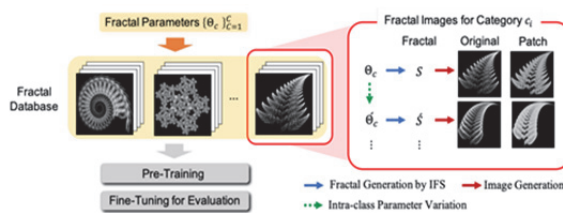


図 3.3-1 FractalDB 構築の概要図 文献[2]より引用

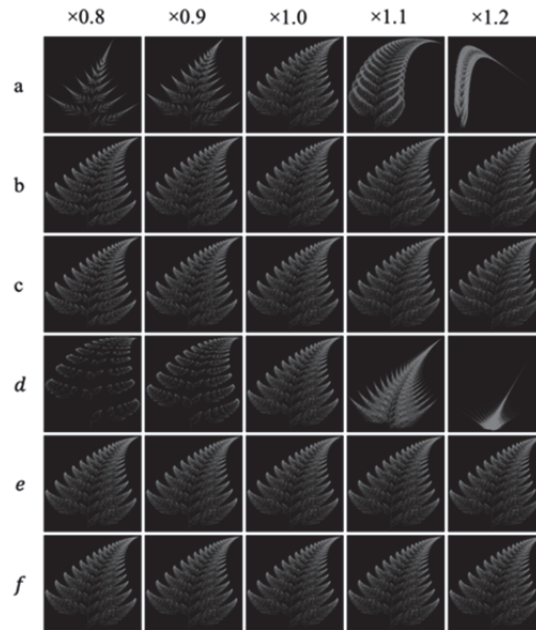


図 3.3-2 アフィン変換のパラメータに摂動を与えた際の形状変 文献[2]より引用

それにより生成される画像のパターンの一部には植物や海岸線といった自然物として認識できる形状ができあがる。このことから、フラクタルは多くの自然物の形状に含まれているとされている。その性質を CNN に直接学習することで、自然画像の認識を向上させることができると仮定して FractalDB は構築された。FractalDB 作成の概略を図 3.3-1 に示す。フラクタル画像の作成には確率付き反復関数系 (IFS; Iterated Function System) を使用した。IFS は完備距離空間 X で定義され、次の式で表される。

$$IFS = \{\chi; \omega_1, \omega_2, \dots, \omega_N; p_1, p_2, \dots, p_N\} \# \quad (1)$$

ここで $\omega_i: \chi \rightarrow \chi$ は変換関数、 p_i は合計が 1 となる確率、 N は関数の数である。IFS を使うことでフラクタル $S = \{x_t\}_{t=0}^{\infty} \in \chi$ はランダムに反復するアルゴリズムによって生成される。このアルゴリズムは初期点 x_0 から $t = 0, 1, 2, \dots$ に対して次の 2 ステップを繰り返す ; (1) i 番目の変換を決定するために、事前に定義した確率 $p_i = p(\omega^* = \omega_i)$ に基づき、 $\{\omega_1, \omega_2, \dots, \omega_N\}$ から使用する変換 ω^* を選択する。(2) 新しい点 $x_{t+1} = \omega^*(x_t)$ を打つ。

画像を作成するため、FractalDB を作成する際の距離空間 χ は 2 次元ユークリッド空間 R^2 とする。また、変換関数には式 (2) のようなアフィン変換を想定している。アフィン変換は回転や移動といった変換を 6 つのパラメータ $\theta_i = (a_i, b_i, c_i, d_i, e_i, f_i)$ で行う。フラクタル画像 S は IFS と ω_i とを用いてドット x を繰り返し打つことによって表現される。

$$\omega_i(x; \theta_i) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} x + \begin{bmatrix} e_i \\ f_i \end{bmatrix} \# \quad (2)$$

FractalDB のカテゴリは ω_i のパラメータ $a_i \sim f_i$ から構成されている。IFS によって生成される画像はパラメータとそれに対応する確率の組み合わせ $\Theta = \{(\theta_i, p_i)\}_{i=1}^N$ で表現が変化する。

FractalDB では、フラクタルのカテゴリは Θ によって固定されると仮定し、ランダムにパラメータをサーチした 1,000 / 10,000 カテゴリを持つデータベースを提案した。FractalDB-1k / 10k は異なる 1,000 / 10,000 個のフラクタルで構成されている。そのフラクタルのパラメータはランダムに探索を行うプログラムを用いて自動的に生成される。パラメータサーチの手順は、まず、 $N = \{2,3,4,5,6,7,8\}$ から N を選択し、次にアフィン変換のパラメータ θ_i を $[-1,1]^6$ の範囲から N 個分決定する。さらに確率 p_i を以下のように設定する。

$$p_i = \frac{(\det A_i)}{(\sum_{i=1}^N \det A_i)} \# \quad (3)$$

$$A_i = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix}$$

最後に、ここまでで設定した $\Theta_i = \{(\theta_i, p_i)\}_{i=1}^N$ を用いてフラクタル画像 S を生成し、その画像の全体に対する点の割合である画面占有率 r が閾値を超えた時に、その Θ をデータセットのカテゴリとして採用した。このようにして図 3.3-2 に示すような画像とそのラベルを持つデータセットを作成した。

FractalDB で生成した画像は黒背景にグレーの点をプロットしたもので表現されている。各カテゴリのインスタンス拡張は、変換関数のパラメータセット Θ に対して $\{0.8 \sim 1.2\}$ かつ 0.1 刻みで摂動を加えることで、フラクタル形状に変化を持たせた。パラメータセット Θ の各パラメータに対して摂動を加え場合のフラクタル形状の変化を図 3.3-2 に示す。また、形状変化に加えて、レンダリングをドットから 3×3 のパッチに変化させることでインスタンスを拡張させている。さらに、これらの画像に対して上下左右反転の処理を施すことで各カテゴリのインスタンス数を 1,000 に統一させている。

これにより FractalDB は人手によるアノテーション作業を要せずにカテゴリ付き大規模画像データセットの自動構築を実現した。さらに、数式からカテゴリをランダムに生成することができるため、理論的にはカテゴリ数を無制限に増やすことも可能である。

3.3.2 「落ちない/落ちてでも安全」なドローンのためのパターン認識技術

NEDO サイバー・フィジカル研究拠点間連携による『革新的ドローン AI 技術』の研究開発プロジェクトは、ドローンの目視外飛行の信頼性・安全性の向上のために、ディープラーニング等の AI 技術を搭載した「落ちない/落ちてでも安全」なドローンの実現を目的とした研究開発を行ってきた。操縦者が飛行中に実施するトラブルシューティングを AI で再現するため、図 3.3-3 のように人や車両を認識し、その上空を回避する自律運航 AI 技術、ドローン搭載のセンサーやテレメトリ情報から故障や不安定な動作を察知する故障診断 AI 技術、故障と診断されたら安全なエリアへ移動し着陸する緊急着陸 AI 技術で構成される。開発した AI を搭載したドローンを、物流シーンを想定した実証実験を先日実施している[3]。



図 3.3-3 「落ちない落ちて安全」なドローンのための AI 技術

ここでは、ドローンの緊急着陸時における着陸可能なエリアを見つけ出す技術について述べる。故障診断 AI により故障を察知したドローンを緊急着陸させる必要がある。このときに、建物や道路に墜落することなく、空き地などの安全な場所を見つけ出し着陸する必要がある。まず、GPS と地図を照合することにより道路や民有地などを避け、河川敷や公園広場などを安全な場所へ移動する。しかしながら、こういった空き地であっても樹木や構造物など、衝突する危険のあるものは避ける必要がある。そこで、ドローンに搭載されたカメラの映像から、セマンティックセグメンテーション等により安全に着陸可能なエリアを推定する。

セマンティックセグメンテーションは画像ピクセル単位で推論を行うタスクで、大局的情報とピクセルの詳細情報を同時に推論可能な U-Net [4] が有名であり、この派生手法が広く使われている。ドローンの安全着陸を目的としても法線方向の垂直・水平・その他の 3 クラス分類と、距離画像を同時に推定する SafeUAV-Net [6] が提案されている。これを利用するにあたり、公開されているネットワークモデルはヨーロッパの建築物を含む風景を学習しており、日本の風景ではうまく動作しない。また、法線方向推定は不要であることから、安全エリア・非安全エリアの 2 クラス分類を行うネットワークモデルを日本の風景で学習した。結果を図 3.3-4 に示す。

SafeUAV-Net にはパラメータ数の違いで Small と Large の 2 つがあり、Small であればドローンに搭載可能な性能のコンピュータでも動作可能である。精度としては 70~80% 程度の Accuracy で推定が可能であった。

一方でカメラを 1 つだけ使用して距離を推定する単眼深度推定 (Monocular Depth Estimation) は、ドメインの異なるデータを混合して学習する技術や、Vision Transformer などの新しいニューラルネットワーク技術により、近年急速に性能が向上している [6]。図 3.3-5 は模擬市街地を飛行するドローンから撮影した映像から、単眼深度推定を行った結果である。建物から落ちる影や、横断歩道などの路面のペイントの部分に誤りはあるが、平面の推定は比較的良好であることがわかる。

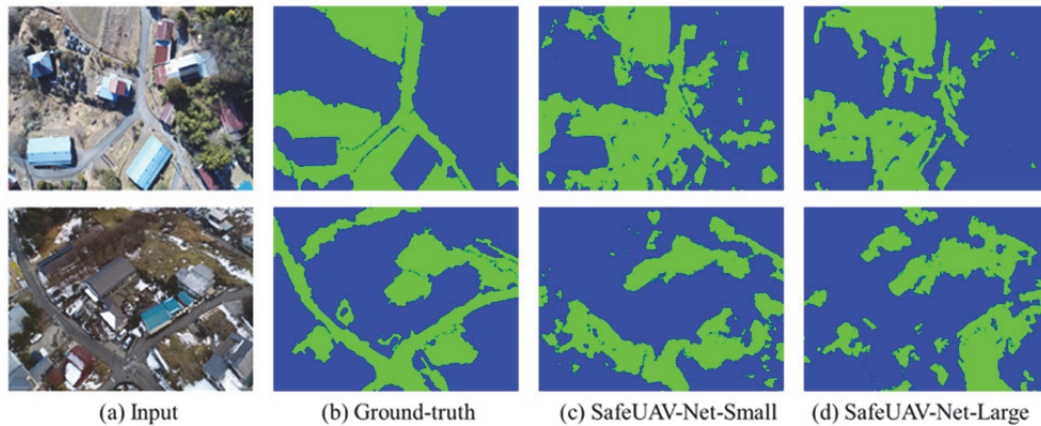


図 3.3-4 SafeUAV-Net による安全エリア・非安全エリア推定



図 3.3-5 単眼深度推定例

3.3.3 物体検出器の精度向上への取り組み

ディープラーニングによる物体検出は、目的とする対象物体の学習データを用意することで、様々な応用が可能である。先述のドローンであれば、物体検出により人や車両を検出し、その上空を回避しながら安全な自律運航を行うことが可能となる。しかし、物体検出器そのものの構造的な制限が存在するため、特に安全を目的として利用する場合には、配慮が必要となる。

物体検出には、特徴抽出と物体検出を一度に行う 1 ステージ型と、物体の候補領域を推定してからその候補領域に対して特徴抽出と物体検出を行う 2 ステージ型が存在する[5]。2 ステージ型の方が一般的に検出精度が高いが、1 ステージ型は処理速度面において有利であるため、機器に組み込むような場合に使いやすい。しかしながら 1 ステージ型物体検出には、検出のグリッドやレイヤーの境界付近で検出精度が低下する問題が指摘されている[7]。図 3.3-6 は 1 ステージ型物体検出として軽量で広く使われている Yolo v1、v2、v3 において、この境界問題を検証した結果である。特定の位置、大きさの物体に対し、検出スコアが極端に低下していることが明確にわかる。ここでは人物 CG を用いて検証しているが、車両などの他の物体でも同様の現象が存在する。

そこで、検出精度が低下するグリッド境界に合わせて高密度にグリッドを配置し、レイヤー境界にも中間サイズのレイヤーを追加することで、精度低下を防ぐ手法を提案している[8]。また、さらなる精度向上を目指し、効率的に学習用データを収集するため、未検出や過検出となりやすい箇所を可視化する方法を提案している[9]。図 3.3-6 は、対象物周辺に背景が映り込むことにより、縦エッジなどの景況により未検出となりやすい箇所を可視化した例である。ドローンでの応

用に限らず、利用シーンを考慮し物体検出器の特性を理解した上で、こういったアプローチにより要求水準を満たす方策が求められていると考える。

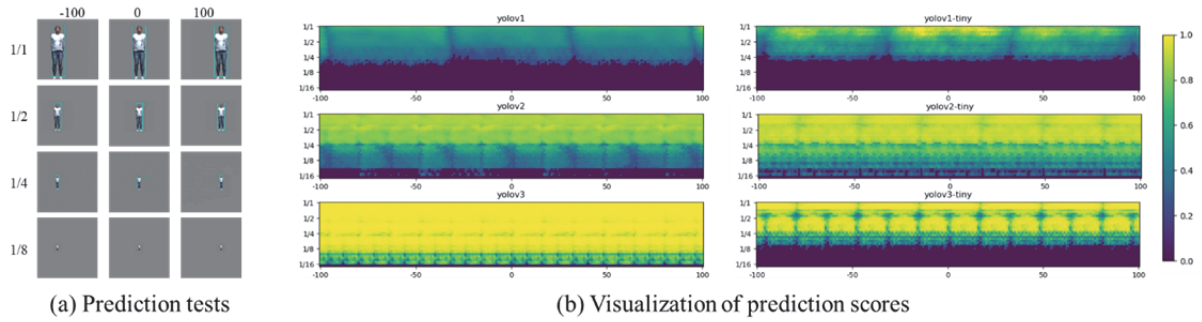


図 3.3-6 物体検出器の境界問題

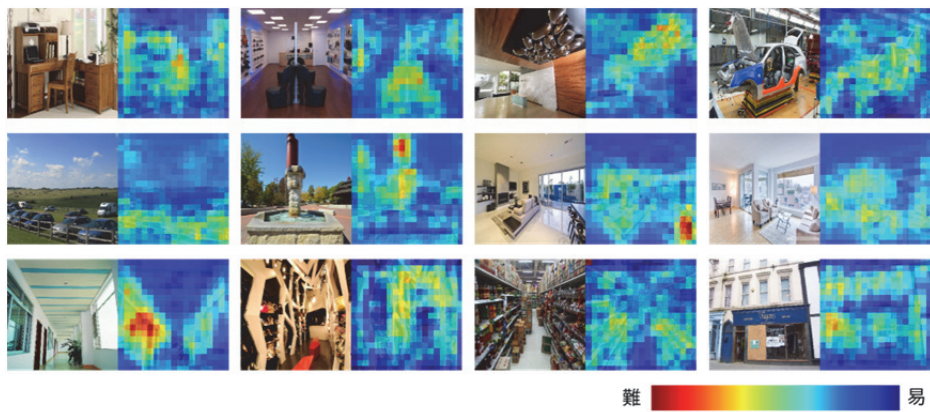


図 3.3-7 未検出となりやすい箇所の可視化

【参考文献】

- [1] 片岡裕雄, 山田亮佑, 松本晟人, 自然の形成原理による CNN の学習, *Medical Imaging Technology*, 2021 年 39 卷 3 号 p. 117-123, <https://doi.org/10.11409/mit.39.117>
- [2] 片岡裕雄, 山田亮佑, 松本晟人, 自然の形成原理による CNN の学習, *Medical Imaging Technology*, 2021 年 39 卷 3 号 p. 117-123, <https://doi.org/10.11409/mit.39.117>
- [3] 「自律運航 AI」を搭載したドローンを用いて荷物配送を行う実証実験を実施, 国立研究開発法人新エネルギー・産業技術総合開発機構プレスリリース, https://www.nedo.go.jp/news/press/AA5_101542.html (2022 年 5 月 18 日) .
- [4] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAD)*, Springer, LNCS, Vol.9351: 234-241, 2015
- [5] Marcu, Alina and Costea, Dragos and Licaret, Vlad and Pirvu, Mihai and Leordeanu, Marius and Slusanschi, Emil. SafeUAV: Learning to estimate depth and safe landing areas for UAVs from synthetic data. *European Conference on Computer Vision (ECCV) UAVision Workshop*. 2018.
- [6] René Ranftl, Alexey Bochkovskiy, Vladlen Koltun, Vision Transformers for Dense Prediction, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12179-12188, 2021.
- [7] A Bochkovskiy, CY Wang, HYM Liao, Yolov4: Optimal speed and accuracy of object detection, *arXiv preprint, arXiv:2004.10934*, 2020
- [8] Yusuke Hosoya, Masanori Suganuma, Takayuki Okatani, Analysis and a Solution of Momentarily Missed Detection for Anchor-based Object Detectors, *WACV2020*, pp. 1399-1407, 2020
- [9] 岩田健司, 松本知浩, 杉本喜一, 物体検出における境界問題の可視化と中間レイヤー・グリッド配置による解決方法, 第 27 回画像センシングシンポジウム SSII2021, SO3-08, 2021
- [10] 岩田健司, 松本知浩, 小林周, 杉本喜一, 物体検出器の学習画像取得回数削減を目的とした未検出・過検出傾向の可視化手法の開発, 第 28 回画像センシングシンポジウム SSII2022, SO3-03, 2022

4. 文字認識システムの技術調査

4.1 OCR の現状

本報告書の 1 章から 3 章には本委員会のミッションやパターン認識技術そして OCR 技術というように様々なレベルや視点から見た現状と未来が展望されているが、この 4.1 節ではより現場に近いところから OCR の現状を俯瞰する。

4.1.1 デジタル化社会時代の OCR

社会の要請は生活に関わるすべての人間の活動の DX 化であり、そのために必要と目されている基礎技術から応用技術、さらに基幹システムからアプリケーション開発には人間を機械とインターフェースする技術が絡んでくる。当然実用化への道程には厳しい現場での評価が待っている。OCR 開発の歴史を紐解くことはパターン認識という極めて複雑で困難な技術開発の悪戦苦闘の歴史をたどることでもあり厳しい現場評価に対応してきた歴史でもある。パターン認識技術の実用化を担ってきた OCR 開発の経験豊かな歴史を読み解けば、温故知新という言葉が古来いわれているとおりに隆盛しつつある AI-OCR を確実に地に根を張っていくための良き栄養源となると思われる。

OCR が熱望される現場はデータ入力業務である。昔からこれは変わっていない。データといっても多様であり本質的でもある。データの質と量は時代の要請に大きく依存する。データ入力業務では効率化が最も表面的な課題として挙げられる。効率化に大きな影響を与える OCR の読み取り結果の品質の向上は社会的な要請レベルの高度化と社会と関わる深度が増すにつれて重要な課題となっていく。質の問題としてまずは正確さがあげられるが現場からの要請をみると AI-OCR といえどもこの課題には万全の回答を用意できていない。その客観的な裏付けは技術的にも難しい。アンケート調査は効率化と品質に関する現場の感覚を探る有力な手段であり、結果の信憑性は不確かではあるが参考文献[18]は一つの見方を示している。また、各社の提供する OCR 広告では使用環境の事前確認の必要性を説いている。それは現場環境の本質に由来する。現実が統計確率的な存在としてあり認知機能はその存在をあるモデルを仮定して最適な解を求める。この原理的な制約は機械認知であろうと人間認知であろうと免れない。したがって、現実の解は絶対ではなくある誤差を許容範囲に収めることで得られる。AI-OCR がこの許容範囲をどれだけ広げたかは各社の実際の読み取り評価によって確かめるしかない。現場それぞれに特定の制限を設けてその範囲内で評価しているのが現状である。有限のリソースで経済的な効果を得るためには OCR 導入前の評価は欠かせない。参考文献[1]はこのあたりの現場の状況がよくまとめられている一例となっている。本報告書にある当委員会でもまとめつつある要因表は現場の声を反映するものとして企画されたもので OCR 導入計画に当たっての有効活用が期待される。

量の問題は参考文献[2]にあるように国家レベル量のアナログ資料のDX戦略にAI-OCRが役立つことを示している。既存のアナログ資料のDXは量を個人レベルで考えても重要である。人間の脳がしばらくはアナログ存在としてありかつ思考の自由さは手書きメモに依存すると仮定するとBMI技術が本格化する未来の話置いて現状は過去のアナログデータのみならず日常生活において発生するアナログデータのDXは重要である。この分野での安価で高性能なAI-OCRが果たす役割は大きい。参考文献[3]以下を見ると古くからの大手メーカーの他に多くのベンチャー企業が数え切れぬほど起業していることが読み取れるが以前に比べればだいぶ現場の要請に答えられる製品へと淘汰整理されて成長しているのがうかがえる。各社それぞれの性能機能を競う様子が見て取れる。残念ながら各社の性能比較基準がないため売れ行きランキング上位から判断するしかないのは昔から変わらない。機能面で見れば世界のIT技術、システムの開発スピードに合わせて使いやすさは年々向上しているがアナログメディアを相手とするにはUIは一層の開発努力が求められている。本報告書にある要因表を活用したOCRに適したアナログデータのキャプチャシステムの開発が求められているといえよう。車の自動運転も難しいがアナログデータの自動採集もまた難しい。キャプチャ機械に人間以上の認知機能性能がいるように思われる。研究開発部門への継続的なかつ強力な投資支援が望まれる。

【参考文献】

- [1] OCRを利用した統計表の体系的なテキストデータ化, 有本寛, 2022年1月, 経済研究, 岩波, https://researchmap.jp/read0150909/published_papers/36310273
- [2] 国会図書館デジタル化, LINE(株)AIカンパニー, 2022年7月, <https://linecorp.com/ja/pr/news/ja/2021/3825>
- [3] 文字認識クラウドサービス, 東芝デジタルソリューションズ, <https://www.global.toshiba.jp/products-solutions/ai-iot/mojigazou.html>
- [4] クラウド型OCR, リコー, <https://www.ricoh.co.jp/service/>
- [5] DynaEye11の発表, PFU, 2022年9月22日, <https://www.pfu.ricoh.com/dynaeye/product/dynaeye11/>
- [6] 最新FormOCR/AI-OCR活用セミナー, メディアドライブ, 2022年11月28日, <https://mediadrive.jp/>
- [7] AI-OCR導入支援ソリューション, NEC, <https://www.nec-solutioninnovators.co.jp/sl/digitallabor/development/aiochr.html>
- [8] 帳票認識サービス, 日立, <https://www.hitachi.co.jp/products/it/bigdata/service/ai-ocr/contents/index.html>
- [9] AI-OCR, パナソニック, <https://www.panasonic.com/jp/business/its/ocr/ai.html>

- [10] DX-OCR, <https://dx-suite.com/>
- [11] データクレンジング・OCR, Double Standard, <https://double-std.com/service/ocr/>
- [12] eas, ㈱うるる BPO, <https://www.uluru-bpo.jp/eas/>
- [13] クラウド型 AI OCR ソリューション CaptureBrain, キヤノン,
<https://canon.jp/business/solution/ocr/lineup/capturebrain>
- [14] 国会図書館デジタルコレクションリニューアル, 2022年12月21日,
https://www.ndl.go.jp/jp/news/fy2022/221221_01.html
- [15] くずし字解読アプリ, 凸版印刷(東京都)と京都市歴史資料館など, 毎日新聞,
2022年12月30日, 「くずし字」解読アプリ1月公開へ 実家に眠る古文書読めるかも,
https://www.toppan.co.jp/news/2022/09/newsrelease220913_1.html
- [16] AI-OCR7 製品まとめて比較, ITトレンド, https://it-trend.jp/ai_ocr
- [17] データ入力がゼロなるクラウド型データ入力サービス WOZE(ウオーゼ),
㈱ハンモック, https://it-trend.jp/ai_ocr/12537
- [18] AI-OCR 導入の実態調査, ペーレス, 2022年6月17日, 【AI OCR 導入の実態調査】
導入後, 約8割が課題あり。うち6割は「確認作業の手間がなくなる」と回答,
https://www.hammock.jp/defact/media/media_pe-less_ocr.html
- [19] AI-OCR 請求書支払い機能強化(SuperStream-NX), Spendia, ITS(株), 2022年10月12日,
Spendia | スーパーストリーム株式会社, <https://superstream.co.jp>

4.2 製品分類について

2023年3月における主要なOCR製品を、一覧表の形で整理した。(表4.2-1～表4.2-20)

以下、一覧表はOCR製品を以下の7つに分類したものとなっている。

- | | |
|------------------------|-----------------|
| ○ ハードOCR製品 - 帳票OCR | 表4.2-1～表4.2-2 |
| ○ ソフトOCR製品 - 帳票OCR | 表4.2-3～表4.2-6 |
| ○ ソフトOCR製品 - 文書OCR | 表4.2-7～表4.2-9 |
| ○ ソフトOCR製品 - 名刺OCR | 表4.2-10 |
| ○ ソフトOCR製品 - 本人確認書類OCR | 表4.2-11～表4.2-14 |
| ○ ソフトOCR製品 - マルチタイプOCR | 表4.2-15 |
| ○ サービスOCR製品 | 表4.2-16～表4.2-20 |

製品分類の視点には、提供形態と対象文書という二つの軸を用いて分類を行った。

第一の整理軸は提供形態に関するものであり、具体的には提供形態がハードウェアかソフト

ウェアかクラウド上のサービスかという区分である。

ハードウェアのOCR製品とは、文字認識における主要な処理を高速で実行するための専用処理装置を備えたものをいう。OCR専用スキャナーと文字認識部とともに同時提供するデバイスタイプと呼ばれる形態が現在では主流となっている。ハードウェアOCRでは一般には通常のスキャナーより高速・高品質なものが用いられることが多く、大量のデータを高速・高精度に処理できることがメリットである。そのメリットを活かして、ハードウェアOCRは定型フォーマットの帳票を読み取り基幹系業務ソフトウェアと連携してデータ処理する用途で用いられることが多い。

一方、CD-ROMやDVDのような情報記録媒体あるいはネットワークを介して、パーソナル・コンピュータ等にインストールして使用するタイプのOCR製品をソフトウェア製品と呼ぶ。スマートフォン、PDA等のモバイル機器にインストールして使用するタイプのOCR製品も含まれる。ソフトウェア実装の利点は、いわゆるモノとしての生産コストが不要であることに加え、画像入力に一般的なスキャナーやデジタルカメラ等を用いることができるため、既存の装置を流用・共用することが可能となり、その結果としてトータルな導入価格を低く抑えうることにある。

近年はユーザーのサーバーやパーソナル・コンピュータにインストールすることを不要とした、クラウド上でOCR処理を行うことで処理結果を得るサービスとしてのOCRも普及してきており、3つ目の提供形態としてサービスOCRという分類を定義した。サービスOCRは導入コストがソフトウェアOCRよりも低価格に抑える代わりに、月額基本料金や利用量に応じたランニングコストを支払う販売方法をとるものも多い。

このように提供形態の違いは、商品形態の違いを生み出し、市場セグメントや販売チャネルなどに大きな差異をもたらしている。したがって、この整理軸は事業的な観点から特に重要な軸と考えることができる。

第二の整理軸は対象原稿に関するものであり、具体的には対象原稿が帳票であるか、一般文書であるか、名刺であるか、本人確認書類（免許証保険、保険証、マイナンバーカードなど）であるかという分類である。今回から新たに複数種類の対象原稿をユーザーの指定なしに対応できるマルチタイプという分類も定義した。

この整理軸は技術的な観点からより重要となる区分である。すなわち、対象原稿が異なることによって、読み取られる文字列の位置や文字種に関する制約に関して大きな差異が生じる。そして、まさに制約こそが情報処理（特にパターン認識系）の実用性（精度・性能）を左右する本質的要素の一つなのである。

帳票OCRが処理対象とする“帳票”は、一般には罫線によって、抽出されるべき文字領域が区分されている。あるいは罫線に代わるものとして、背景パターン・色や並びの整然さ等を仮定することができる。対象文字は、頁全体としては手書き及び印刷された活字の両方を含むことが可能だが、個々の記入領域に関しては制約が強いのが一般的である。制約は文字の種類によるもの

だけでなく、意味（部品名か数量か、住所か氏名かなど）によるものも含む。そのため、辞書的な知識を適用することで、実用性（精度・性能）を向上させることが可能となっている。帳票OCRの主な用途として、受発注業務やアンケート集計におけるデータ・エントリーがあげられる。

名刺OCRが処理対象とする“名刺”は、基本の大きさは55mm×91mmの限られた小さな紙に、氏名、会社名、住所、電話番号、メールアドレスなどの項目を記したものである。一般文書とは異なり、記載されている項目は限定され、ある一定のレイアウトデザインが存在する。名刺の向きを自動的に補正し、レイアウト解析と呼ばれる処理によって、項目の異なるそれぞれの領域を抽出する必要があるが、大容量の電話番号辞書や郵便番号辞書を搭載し、会社名や住所の誤認識を自動修正することにより、十分な実用性をもたせることが可能となっている。名刺OCRの主な用途として、登録された名刺の検索、閲覧機能だけでなく、登録された住所の地図を調べる機能、登録された住所までの経路を調べる機能、個人情報保護法の施行に伴い、名刺情報を保護するセキュリティ機能まで搭載された商品が発売されている。

本人確認書類は、2015年10月に施行されたマイナンバー制度に関連する書類、あるいは運転免許証などの本人確認書類を対象とするOCRソフトウェアを想定している。各社今後こういった認識のソリューションが増えることを想定して専用にパッケージングしていると考えられる。

文書OCRが処理対象とする“文書”は、広くは帳票、名刺、本人確認書類ではないその他一般文書という意味である。具体的には、書籍・新聞・雑誌・論文・報告書・通達文などがこれに該当する。（設計図面や建築図面などは該当しないとされる。）文書の構成には、罫線のような明確なセパレータが用いられないため、レイアウト解析によって、属性の異なるそれぞれの領域を抽出する必要がある。文字の種類に関する制約はほとんどないといってよく、わずかに「特定言語で書かれた」「印刷（活字）文字」であるという程度である。しかしながら、この制約もそれなりに強力なものであって、十分な実用性をもたせることが可能となっている。文書OCRの主な用途として、印刷文書の再利用（テキスト化）、及び検索性の付与が上げられる。法人市場のみならず個人市場においても利用可能である点が、帳票OCRとは異なっている。

このように、事業的な観点から重要な提供形態という整理軸と、技術的な観点から重要な対象原稿という整理軸との二つの軸を使用して整理を行った。ただし近年、ハードウェアOCRは帳票分野以外存在しないので、ソフトウェアOCRとサービスOCRの2つが対象原稿の軸で分類される形になっている。それぞれの分類の動向を把握することにより、OCR市場の概略を事業的観点及び技術的観点から理解することが可能となるだろう。

表4.2-1 ハードOCR製品(帳票OCR装置) その(1)

製品名	メーカー	処理速度 (枚/分)		文字 認識 速度 (/s)	スキヤナ					機能				規格部 HxDxW mm Kg	インター フェース	フォーマット 指定方式	発売年月	価格(万円) 【税別】
		A4 300字	A8 10字		帳票 サイズ	ホツパ 容量	アケセ メント	スタツカ容 量	リジエ クト	解像度 (DPI)	帳票厚 (運量)	漢字読取 手書/活字	知識 処理					
HT-4161	日立	約60 (アソソク 使用時約 85)	約120		52x74~ 305x420	約300枚	約300枚	約100(2 スタツカ容 量のみのみ)	100~600	35kg~ 180kg紙	OP手書/活字 約4,300種	OP	あり	低電力モード(スリー プモード), OP 異サイ ズ帳票混在/画面読取/ カラー画像入力	機構部 HxDxW mm Kg 1スタツカ容 約351x約331x約 410(ホツパ・スタツカ取 納時) 約351x約353x約 410(ホツパ・スタツカ使 用時) 約18.5kg	回路部 HxDxW mm Kg (機構部を含む)	2012/7	1スタツカ容 #:131万円~ 2スタツカ容 #:181万円~
HT-4165	日立	約210	約380		52x74~ 305x458	約550枚	約550枚	約100枚	100~600	35kg~ 180kg紙	OP手書/活字 約4,300種	OP	あり	低電力モード(スリー プモード), OP 異サイ ズ帳票混在/画面読 取/カラー画像入力	機構部 HxDxW mm Kg 約407x約715x約 485 約638x約980x約 485(延長トレイ使 用時) 約43kg	回路部 HxDxW mm Kg (機構部を含む)	2015/7	片面容 #:241万円~ 両面容 #:331万円~
Blinkscan BS300	日立チャネ ルソリューションズ株 式会社	—	—	—	216mm× 297mm	—	—	—	200/240/3 00/400/48 0	複写界深 度10mm	—	—	あり	カラー読取り	機構部 HxDxW mm Kg (機構部を含む)	—	2011/4	オープン 価格

表4.2-2 ハードOCR製品(帳票OCR装置) その(2)

製品名	メーカー	処理速度 (枚/分)		文字 認識 速度 (/s)	スキヤナ					機能				機構部 HxDxW mm Kg	回路部 HxDxW mm Kg	インター フェース	フォーマット 指定方式	発売年月	価格(万 円) [税別]
		A4 300字	A8 10字		帳票 サイズ	ポッパ 容量	スタック容量 シート リジエ クト	解像度 (DPI)	帳票厚 (連量)	漢字読取 手書/活字	知識 処理	画像 出力	その他 機能						
N6370E	日本電気	120/ 210	300/ 420		74x52~ 420x305	55mm	55mm	10mm	100~600	40.7~ 209.3/m ²	OP手書約 4000種	あり	あり	あり	435x545x450 49kg	USB3.0	フォーマット プログラム	2017/9	160~
N6370M	日本電気	55	—		105x75~ 297x235	16mm	16mm	10mm	100~600	40.7~ 174.5g/ m ²	OP手書約 4000種	あり	あり	あり	275x420x380 24kg	USB3.0	フォーマット プログラム	2018/2	88~
OCR2000i (モデル2700)	東芝デジタル ソリューションズ	120	—		70x52~ 420x305	35mm	35mm	10mm	200/240/ 300/ 400/600	35kg~ 180kg	OP手書・OP活 字4000種	OP	あり	あり	350x600x450 約30kg	USB3.0	GUIによる帳 票定義	2014/3	255~
OCR2000i (モデル3500)	東芝デジタル ソリューションズ	220	—		70x52~ 458x305 (オブ ションで 70x52~ 630x305)に も対応)	80mm	80mm	25mm	100/150/ 200/240/ 300/400/ 600	35kg~ 180kg	OP手書・OP活 字4000種	OP	あり	あり	498x700x430 約75kg	USB3.0	GUIによる帳 票定義	2012/10	305~
OCR2000i (モデル12000)	東芝デジタル ソリューションズ	330	—		70x52~ 420x305 (80x40~ 260x52)に も対応)	150mm	150mm	50mm× 3	200/240/ 300/400	35kg~ 180kg	OP手書・OP活 字4000種	OP	あり	あり	1325x740x19 25 420kg	専用	GUIによる帳 票定義	2016/11	2,685~
S2700EH	東芝デジタル ソリューションズ	120	—		70x52~ 420x305	35mm	35mm	10mm	200/240/ 300/ 400/600	35kg~ 180kg	OP手書・OP活 字4000種	OP	あり	あり	350x600x450 約30kg	USB3.0	GUIによる帳 票定義	2019/3	292~
S3500Z	東芝デジタル ソリューションズ	220	—		70x52~ 458x305 (オブ ションで 70x52~ 630x305)に も対応)	80mm	80mm	25mm	100/150/ 200/240/ 300/400/ 600	35kg~ 180kg	OP手書・OP活 字4000種	OP	あり	あり	498x700x430 約75kg	USB3.0	GUIによる帳 票定義	2021/10	451~

表4.2-3 ソフトOCR製品(帳票OCR その1)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
帳票OCR サーパー版 Ver.2.02	パナソニック ソリューション テクノロジ	手書き：英字、数字、ひらが なの一部、カタカナの一部、 記号の一部、第1水準漢字、第 2水準漢字の一部(518字) 活字：約6800字、数字、英 大、カナ、ひらがな、JIS記号 (一部)168字、ギリシヤ文字 (一部)32字、第1水準漢字、 第2水準漢字	手書き、活字、バーコー ド(NW-7/CODE39/GSI- 128/CODE128/JAN-8/JAN- 13/ITF-6/ITF-14/ITF-16)、QRコード、チェック マーク(レ点、塗りつぶ し、オーバーライト)	6~60ポイント (400dpiの場 合)		Windows Server 2016 / 2012 R2/ 2012 / 2008 R2 SP1以上/ 2008 SP2以上 いずれも 日本語版	お使いのOS が推奨する 環境以上	2017/2	スーパー版 2,000,000 [税別]	
AI帳票OCR Ver.9 (WisOCR)	パナソニック ソリューション テクノロジ	手書き：英字、数字、ひらが なの一部、カタカナの一部、 記号の一部、第1水準漢字、第 2水準漢字の一部 活字：約6800字、数字、英 大、カナ、ひらがな、JIS記号 (一部)168字、ギリシヤ文字 (一部)32字、第1水準漢字、 第2水準漢字	AI手書き、マルチフォン ト(明朝体、ゴシック 体、教科書体、ワープロ 体、新聞文字など)、 バーコード、QRコード、 チェックマーク	6~60ポイント (400dpiの場 合)		Windows 11 version 21H2/22H2 Windows 10 version 21H2/22H2	お使いのOS が推奨する 環境以上		要問合せ [税別]	Basic/Standard/Pro 月 額基本利用料 PC上で動作するクライ アントアプリとクラウ ドのAI-OCRエンジンで 構成される。
帳票認識ライブラリー Ver.8.51	パナソニック ソリューション テクノロジ	手書き(英字、数字、ひらが なの一部、カタカナの一部、 記号の一部、第1水準漢字、第 2水準漢字の一部(518字))、 活字(約6800字、数字、英 字、カナ、ひらがな、JIS記号 (一部)168字、ギリシヤ文 字(一部)32字、JIS第1水 準漢字、JIS第2水準漢字)	手書き、活字、バーコー ド(NW-7/CODE39/GSI- 128/CODE128/JAN-8/JAN- 13/ITF-6/ITF-14/ITF-16)、チェックマーク(レ 点、塗りつぶし、オー バーライト)	6~60ポイント (400dpiの場 合)		Windows 10/8.1/Windows Server 2016/2012 R2/2012	対応OSが必 要とする最 低メモリ一 に加えて 256MB以上 (512MB以上 を推奨)	2021/2	550,000 [税別]	AI手書き文字認識オペ ションあり(価格は別 途ご相談)
FAX/OCR SYSTEM 伝匠 V9	リコーージャパン	手書き(数字、英大文字、カ ナ、漢字、ひらがな、記 号)、活字(数字、英字、カ ナ、漢字、ひらがな、記号)	手書き、明朝、ゴシッ ク、(OCR-B, Original)	手書き(5mm以 上推奨)、活字 (6~60ポイン ト推奨)		Windows 10/11 Server/2012R2/201 6	OCRサーバ： 1GB以上 結果修正： 512MB以上推 奨	2012/04	680,000~ [税別]	FAXサブシステム無。自 動方向判別機能。自動 帳票識別機能。手書き 住所辞書、氏名辞書標 準添付。帳票定義ソー ル付き
FAX/OCR SYSTEM 伝匠 V10	リコーージャパン	手書き(数字、英大文字、カ ナ、漢字、ひらがな、記 号)、活字(数字、英字、カ ナ、漢字、ひらがな、記号)	手書き、明朝、ゴシッ ク、(OCR-B, Original)	手書き(5mm以 上推奨)、活字 (6~60ポイン ト推奨)		OCR処理側： Windows Server/2016/2019 認識結果修正側： Edge/Chrome	サーバ：8GB 以上 結果修正： 2GB以上	2017/12	880,000~ [税別]	バーコード、QRコード の認識。TIFFに加え て、カラー画像、PDF ファイルも処理可能。 OCR結果確認修正をWeb ブラウザを使用(クラ ウド配信不要)。

表4.2-4 ソフトOCR製品(帳票OCR その2)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
WinReader Hand S v. 6.0	NTTデータ NJK (メデイアアド ライブ)	手書(英数字,カタカナ, ひらがな,漢字4419文字, 記号)	手書(OCR-HN, OCR- HK, OCR-HS, OCR- B, OCR-Kなど)	4x6mm~ 12x12mm		Windows 11/10/8.1	2GB以上	2013/3	300,000円 [税別]	AI-OCR機能(フリーペッチ手 書き文字認識機能、他搭載) ServerOS(Windows Server2022/2019/2016) 対応 版あり、色指2値化オブショ ン、カラー分離帳票認識オブ ションあり
FormOCR v. 8.0	NTTデータ NJK (メデイアアド ライブ)	手書(英数字,カタカナ, ひらがな,漢字4419文字, 記号) 活字(英数字,カタカナ, ひらがな,漢字6355文字, 記号) バーコード/QRコード	手書(OCR-HN, OCR- HK, OCR-HS, OCR- B, OCR-Kなど) 活字(マルチフォ ント, IBM407, 12F)	手書(4x6mm~ 12x12mm) 活字(3~ 15mm)		Windows 11/10	8GB以上	2022/8	600,000円 [税別]	AI-OCR機能(フリーペッチ手 書き文字認識機能、他搭載) ServerOS(Windows Server2022/2019/2016/2012R 2/2012) 対応版あり、色指2 値化オブション、カラー分離 帳票認識オブションあり
帳票認識ライブ ラリ v. 9.5	NTTデータ NJK (メデイアアド ライブ)	手書(英数字,カタカナ, ひらがな,漢字4419文字, 記号) 活字(英数字,カタカナ, ひらがな,漢字6355文字, 記号) バーコード/QRコード	手書(OCR-HN, OCR- HK, OCR-HS, OCR- B, OCR-Kなど) 活字(マルチフォ ント, IBM407, 12F)	手書(4x6mm~ 12x12mm) 活字(3~ 15mm)		Windows 11/10/8.1	2GB以上	2021/6	1,500,000円 [税別]	AI-OCR機能(フリーペッチ手 書き文字認識機能、他搭載) ServerOS(Windows Server2022/2019/2016/2012R 2/2012) 対応版あり、色指2 値化オブション、カラー分離 帳票認識オブションあり
OCR2000i	東芝デジタ ルソリューションズ	数字、英字、カナ、記 号、漢字 バーコード(JAN, ITF, NW-7, CODE39, CODE128, GS1-128) 2次元コード(QRコード) タイミンクマーク	手書き 活字(JIS OCR-B, OCR-K, IBM407, マ ルチフォント, オ ムニフォント, 7B, E13B, 12F)			Windows 10/8.1/7	1GB以上	2013/8	1,050,000円~ [税別]	
OCR2000i モデル LE	東芝デジタ ルソリューションズ	数字、英字、カナ、記 号、漢字 バーコード(JAN, ITF, NW-7, CODE39, CODE128, GS1-128)、2次元コード (QRコード)	手書き、活字(JIS OCR-B, OCR-K, IBM407, マルチ フォント, オムニ フォント, 7B, E13B, 12F)			Windows 10/8.1/7	1GB以上	2016/4	420,000円~ [税別]	

表4.2-5 ソフトOCR製品(帳票OCR その3)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
DynaEye 10	PFU	活字、手書き文字(数字、日本語)、手書きマーク、QRコード、バーコード	手書き、OCR-B, OCR-K, MS明朝, MSゴシック, 手書きフリービット(オプション)			Windows 7/8.1/10/11, Windows Server 2022	2GB以上 (4GB推奨)	2019/09	要問合せ	標準アプリケーション、帳票OCR Entry、SDK/ランタイムなどの提供形態あり
DynaEye 11	PFU	活字、手書き文字、フリービット手書き文字、QRコード、バーコード	手書き、活字(オムニフォント), 手書きフリービット			Windows 7/8.1/10/11, Windows Server 2022 (AMR版除く)	2GB以上 (4GB推奨) AI-OCRは8GB以上(12GB推奨)	2022/07	Entry:初期 1,008,000円・継続 168,000円/年 AI-OCR:初期 2,016,000円・継続 336,000円/年等	DynaEye 11 Entry、DynaEye 11 Entry AI-OCR、DynaEye 11 Entry マルチステーション、SDK/ランタイムなどの提供形態あり
FUJITSU AI-OCR 「Keyword Capture Client Edition」	富士通プロ ンテック					Windows 10 (64bit)	4GB以上(推奨8GB以上)	2019/10	個別見積	帳票定義を用いる「定義型」と、キーワード登録により汎用帳票に対応する「定義レス型」の両方のタイプの帳票認識機能を提供
imageWARE Scan Manager DS V1.1	キャノン	手書き(数字、英大文字、カナ、ひらがな、漢字、記号)、活字(日本語、英語)	手書き、OCR-B、MICR、チェックラ イター文字、明朝、 ゴシック、 Helvetica, Courier New, Times New Roman	6ポイント～ 48ポイント 相当		Windows XP/Vista/7/8 Windows Server 2008/2012	1GB以上	2012/5	400,000 [税別]	
MELFOS	三菱電機	—	—	—	—	—	—		要問合せ	サービスも有

表4.2-6 ソフトOCR製品(帳票OCR その4)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
GAZOMAGIC	日立チャネルソリューションズ	—	—	—	—	—	—	—	要問合せ	TWAIN対応、バーコード認識、印鑑照合、帳票識別機能等。住所、氏名知識辞書、帳票定義ユーティリティあり。手書きフリーピッチ(住所、氏名,数字列等)
帳票マスタ SE V4	日立ソリューションズ・テクノロジーズ	活字、手書き	英数字、ひらがな、カタカナ、JIS第1水準漢字、JIS第2水準漢字、記号			MS Windows/MS Windows Server 詳細は要問合せ	1GB以上 (3GB推奨)		580,000 [税別]	
帳票マスタ LE V4	日立ソリューションズ・テクノロジーズ	活字、手書き	英数字、ひらがな、カタカナ、JIS第1水準漢字、JIS第2水準漢字、記号			Windows7 (32/64ビット版)	1GB以上 (3GB推奨)		380,000 [税別]	
TeLeForm	Hammock	—	—	—	—	—	—	—	要問合せ	
AnyForm OCR	Hammock							2017/3/1	5,300,000円 [税別]～ ※製品構成により価格が異なります。	
AI Read	アライズイノベーション	活字、手書き	—	—	—	—	—	—	要問合せ	事例：注文書、請求書、決算書、アンケート用紙、チェックリストア共通伝票、特許関連書類

表4.2-7 ソフトOCR製品(文書OCR その1)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	TWAIN対応	メモリ	発売年月	価格(円)	備考
活字文書OCRライブラリ v.9.5	NTTデータNJK (メデイアアド ライブ)	約6700+英語+中韓 語<以下オプシ ョン> タイ語+ベトナム 語+マレー語+イ ンドネシア語	明朝、ゴシック、 教科書体などのマ ルチフォント	5~45ポイント		Windows 11/10/8.1			2017/7	500,000円 [税別]	ServerOS (Windows Server 2022/2019/2016/2012R 2/2012) 対応版あり
活字文書OCRライブラリ for Linux	NTTデータNJK (メデイアアド ライブ)	約6700+英語	明朝、ゴシック、 教科書体などのマ ルチフォント	5~45ポイント		CentOS 7 Red Hat Enterprise Linux 7/8			2012/8	550,000円 [税別]	中韓語認識オプシ ョン あり
活字文書OCRライブラリ for iOS/Android	NTTデータNJK (メデイアアド ライブ)	約3700 (英語含む)	明朝、ゴシック、 教科書体などのマ ルチフォント	7.5~45ポイン ト		iOS 15~16 Android 9~13			2012/6	550,000円 [税別]	中韓語認識オプシ ョン あり オプシ ョン認識対象文 字種: 中国語(簡体字) 6763 中国語(繁体字) 13053 韓国語Hangul 2350 韓国語Hanja 4888
OCRパッケージ4	日立	-	-	-	-	-	-			要問合せ	
Mobile 0mCR	オムロン ソ フトウェア	辞書サイズによる	辞書サイズによる ドット文字OCR有 り			32bit/64bit CPU マルチOS対応 (Android™, iOS Linux, Symbian, WindowsMobile , etc.)				要問合せ	

表4.2-9 ソフトOCR製品(文書OCR その3)

製品名	メーカー	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円) [税別]	備考
e. Typist v. 15.0	NTTデータNJK (メデイアアド ライズ)	約6700+欧米55 言語+中韓語	明朝、ゴシック、教科 書体などのマルチフォ ント	5~45ポイント		Windows 11/10/8.1	○	1GB以上	2013/9	19,800円 [税別]	
WinReader PRO v. 15.0	NTTデータNJK (メデイアアド ライズ)	約6700+欧米55 言語+中韓語	明朝、ゴシック、教科 書体などのマルチフォ ント	5~45ポイント		Windows 11/10/8.1	○	2GB以上	2014/11	198,000円 [税別]	ServerOS(Windows Server 2022/2019/2016/2012 R2/2012) 対応版あ り、OLEオートメー ション開発キットオ プションあり
ドキュメント リダー ExpressReader Pro V4.5	東芝デジタル ソリューションズ	約4000	オムニフォント	6~40ポイント	1,200以上	Windows10 x64(1909) Enterprise/ Standard	○	512MB以上		195,000円	
Rosetta-Stone- Components V1.71	キヤノンマー ケティング ジャパン	手書き(教 字、英大文 字、カナ、ひ らがな、漢 字、記号)、 活字(日本語、 英語)	手書き、OCR-B、M I C R、チエックライター 文字、明朝、ゴシッ ク、Helvetica、 Courier New、Times New Roman	6ポイント~48 ポイント相当		Windows 10 Windows Server 2012 R2/2016		256MB以上推 奨	2016/10	オープン	前処理(適応的二値 化、方向判別他)、 帳票認識/登録、印影 抽出、チエックマー ク、丸囲み判定、固 有名詞、住所、氏名 知識辞書あり。

表4.2-10 ソフトOCR製品(名刺OCR その1)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	TWAIN対応	メモリ	発売年月	価格(円)	備考
名刺認識ライブラリー Ver. 3.20	パナソニックソリューションテクノロジーズ	約6,800字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6~60ポイント(400dpi)	—	Windows 10/8.1/7/Windows Server 2016/2012 R2/2012/2008 R2 SP1/2008 SP1 SP2		128MB(推奨256MB)以上	2018/2	550,000 [税別]	
名刺認識ライブラリー Ver. 3.10 for iOS / Android	パナソニックソリューションテクノロジーズ	約6,800字	マルチフォント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6~60ポイント(400dpi)	—	iOS 6.0 / 6.1 / 7.0 / 7.1 / 8.0 / 8.1 Android 4.0 / 4.1 / 4.2 / 4.3 / 4.4		OSが必要とする最低メモリーに加えて128MB以上(256MB以上を推奨)	2015/3	550,000 [税別]	
本格読取おまかせ名刺管理3	ソースネクスト	—	—	—	—	Windows 10 (32ビット/64ビット版)	○	1GB以上	2016/11	ダウンロード版 2,178(税込)	
名刺認識ライブラリー v. 8.0	NTTデータNJK(メデライアドライブ)	約3700+英語+中韓言語	明朝、ゴシックほかマルチフォント	5~36ポイント		Windows 11/10/8.1 Windows Server 2022/2019/2016/2012R2/2012			2016/1	500,000 [税別]	名刺レイト自動解析、法人電話番号辞書搭載
名刺認識ライブラリー v. 3.0 for Linux	NTTデータNJK(メデライアドライブ)	約3700+英語+中韓言語	明朝、ゴシックほかマルチフォント	5~36ポイント		CentOS 7 Red Hat Enterprise Linux 6/7/8			2015/9	500,000 [税別]	名刺レイト自動解析
やさしく名刺ファインディングPRO v. 15.0	NTTデータNJK(メデライアドライブ)	約3700+英語+中韓言語	明朝、ゴシックほかマルチフォント	5~36ポイント		Windows 11/10/8.1	○	1GB以上	2017/11	7,800 [税別]	

表4.2-II ソフトOCR製品(本人確認書類 その1)

製品名	メーカー	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
免許証認識ラ イブラリ v.5.0	NTTデータ (メデイアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(一() [] ノ・ター)	明朝体、ゴシツ ク体、他マルチ フォント対応	42.5 ~ 236 ピクセル 角程度 (500万画素カメラで カード全体がちょうど 収まるサイズの場合、 1.8 ~ 10 mm 角程 度)		Windows 11/10/8.1	—		2020/5/12	300,000円 [税別]	ServerOS(Windows Server2022/2019 /2016/2012R2/20 12) 対応版あり
免許証認識ラ イブラリ v.6.0 for Linux	NTTデータ (メデイアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(一() [] ノ・ター)	明朝体、ゴシツ ク体、他マルチ フォント対応	42.5 ~ 236 ピクセル 角程度 (500万画素カメラで カード全体がちょうど 収まるサイズの場合、 1.8 ~ 10 mm 角程 度)		CentOS 7 Red Hat Enterprise Linux 7/8	—		2021/6/18	900,000円 [税別]	
免許証認識ラ イブラリ v.5.0 for iOS	NTTデータ (メデイアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(一() [] ノ・ター)	明朝体、ゴシツ ク体、他マルチ フォント対応	42.5 ~ 236 ピクセル 角程度 (500万画素カメラで カード全体がちょうど 収まるサイズの場合、 1.8 ~ 10 mm 角程 度)		iOS 15~16	—		2020/5/12	500,000円 [税別]	
免許証認識ラ イブラリ v.5.0 for Android	NTTデータ (メデイアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(一() [] ノ・ター)	明朝体、ゴシツ ク体、他マルチ フォント対応	42.5 ~ 236 ピクセル 角程度 (500万画素カメラで カード全体がちょうど 収まるサイズの場合、 1.8 ~ 10 mm 角程 度)		Android 9~13	—		2020/5/12	500,000円 [税別]	
在留カードOCR ライブラリ for Linux	NTTデータ (メデイアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(一()・タノ)	明朝体、ゴシツ ク体、他マルチ フォント対応	42.5 ~ 236 ピクセル 角程度 (500万画素カメラで カード全体がちょうど 収まるサイズの場合、 1.8 ~ 10 mm 角程 度)		CentOS 7 Red Hat Enterprise Linux 7/8	—		2021/6/18	900,000円 [税別]	

表4.2-12 ソフトOCR製品(本人確認書類 その2)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ(400DPI)	認識速度(字/秒)	OS	TWAIN対応	メモリ	発売年月	価格(円)	備考
マイナンバーカードOCRライブラリ	NTTデータNJK (メデイアアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(ー()・々 /ー)	明朝体、ゴシック 体、他マルチ フォント対応	42.5 ~ 236 ピク セル角程度 (500万画素カメラ でカード全体が ちようど収まるサ イズの場合、1.8 ~ 10 mm 角程度)		Windows 11/10/8.1	—		2015/11/2	500,000円 [税別]	ServerOS (Windows Server 2022/2019/2016/2 012R2/2012) 対応 版あり
マイナンバーカードOCRライ ブラリ v.2.0 for Linux	NTTデータNJK (メデイアアド ライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(ー()・々 /ー)	明朝体、ゴシック 体、他マルチ フォント対応	42.5 ~ 236 ピク セル角程度 (500万画素カメラ でカード全体が ちようど収まるサ イズの場合、1.8 ~ 10 mm 角程度)		CentOS 7 Red Hat Enterprise Linux 7/8	—		2016/5/30	900,000円 [税別]	
DynaEye 運転免許 証OCR	PFU					Windows 10/8.1/8/7/Vista/ XP	—	Windows 7以 上: 32bitOS:1GB 以上 64bitOS:2GB 以上 Windows Vista/XP: 512MB以上	2015/12	202,000円 [税別]	A6コンパクトフ ラットベツドス キャナfi-65F/fi- 60Fを使用
DynaEye マイナ ンバーOCR	PFU					Windows 10/8.1/8/7	—	32bitOS:1GB 以上 64bitOS:2GB 以上	2015/12	52,000円 ~ [税別]	対応スキャナ: fi-65F、fi-60F、 fi-800R、 ScanSnap iX100、 ScanSnap S1100 対象カード:通知 カード表面、個人 番号カード表面/ 裏面
DynaEye 本人確 認カメラOCR V5.0	PFU					Windows 8.1/10 Android 7.1.1/10 iOS 8.1~12.4, iPadOS 15	—		2017/12	要相談	Arrows Tab、 Xperia、iPad等、 タブレット端末の カメラで読み取っ た本人確認書類 (運転免許証/マ イナンバーカード /在留カード)を 認識

表4.2-13 ソフトOCR製品(本人確認書類 その3)

製品名	メーカー	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
マイナバンパー カード認識ライ ブラリー Ver.2.0	パナソニック ソリューション テクノロ ジー	約6,800字	—	6~60ポイント (400dpi)	—	Windows 11/10/8.1 Windows Server 2022/2019/2016/2012 R2/2012	—	OSが必要とする 最低メモリーに 加えて256MB以上 (512MB以上を推 奨)	2022/1	550,000円 [税抜]	
マイナバンパー カード認識ライ ブラリー for iOS Ver.2.0	パナソニック ソリューション テクノロ ジー	約6,800字	—	6~60ポイント (400dpi)	—	iOS 13/14/15	—	OSが必要とする 最低メモリーに 加えて256MB以上 (512MB以上を推 奨)	2022/1	550,000円 [税抜]	
マイナバンパー カード認識ライ ブラリー for Android Ver.2.0	パナソニック ソリューション テクノロ ジー	約6,800字	—	6~60ポイント (400dpi)	—	Android 10/11/12	—	OSが必要とする 最低メモリーに 加えて256MB以上 (512MB以上を推 奨)	2022/1	550,000円 [税抜]	
免許証認識ライ ブラリー Ver.3.20	パナソニック ソリューション テクノロ ジー	約6,800字	マルチフオン ト(明朝体、ゴシ ック体、教科書 体、教科書体、 ワープロ体、新 聞文字など)	6~60ポイント (400dpi)	—	Windows 10 version 21H1/21H2 Windows11 version 21H2/22H2 Windows 8.1 SPなし Windows Server 2022 version 21H2/2019 version 1809 /2016 version 1607/ 2012 R2/2012	-	OSが必要とする 最低メモリーに 加えて、256MB以 上 (512MB以上を 推奨)	2021/10	550,000円 [税抜]	
免許証認識ライ ブラリー Ver.3.20 for iOS	パナソニック ソリューション テクノロ ジー	約6,800字	マルチフオン ト(明朝体、ゴシ ック体、教科書 体、教科書体、 ワープロ体、新 聞文字など)	6~60ポイント (400dpi)	—	iOS 14/15/16	—	OSが必要とする 最低メモリーに 加えて128MB以上 (256MB以上を推 奨)	2021/10	550,000円 [税抜]	
免許証認識ライ ブラリー Ver.3.0 for Android	パナソニック ソリューション テクノロ ジー	約6,800字	マルチフオン ト(明朝体、ゴシ ック体、教科書 体、教科書体、 ワープロ体、新 聞文字など)	6~60ポイント (400dpi)	—	Android 9/10/11	—	OSが必要とする 最低メモリーに 加えて128MB以上 (256MB以上を推 奨)	2021/10	550,000円 [税抜]	

表4. 2-14 ソフトOCR製品(本人確認書類 その4)

製品名	メーカー	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
自動車運転免許 解析ライブラリ	アイエスピー			サポート画像サ イズ 写真：推 奨画素数 500万 画素 (カード部 分が300万画素を 要求)		サーバー上での画像解 析及びスマートフォン アプリ内での解析	—			要相談	
マイナンバー カード解析ライ ブラリ	アイエスピー			サポート画像サ イズ 写真：推 奨画素数 500万 画素 (カード部 分が300万画素を 要求)		サーバー上での画像解 析及びスマートフォン アプリ内での解析	—			要相談	
領収書解析ライ ブラリ	アイエスピー			サポート画像サ イズ 写真：推 奨画素数 500万 画素 (カード部 分が300万画素を 要求)		サーバー上での画像解 析及びスマートフォン アプリ内での解析	—			要相談	
名刺解析ライブ ラリ	アイエスピー			サポート画像サ イズ 写真：推 奨画素数 300万 画素～		サーバー上での画像解 析及びスマートフォン アプリ内での解析	—			要相談	
在留カードライ ブラリ	アイエスピー			サポート画像サ イズ 写真：推 奨画素数 300万 画素～		サーバー上での画像解 析及びスマートフォン アプリ内での解析	—			要相談	

表4.2-15 ソフトOCR製品(マルチタイプ)

製品名	メーカー	認識対象 文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	TWAIN 対応	メモリ	発売年月	価格(円)	備考
本人確認書類読 取ソフトウェア	東芝デジタル ソリューションズ	約4000	オムニフォント	6~40ポイント	5秒以下/1枚	Windows Server 2012 R2 Windows Server 2016 RedHat Enterprise Linux 8.3	-	1GB以上	2020/10	月額利用料 金 150万円~ (免許証読 取5万枚。開 発キット、別 メデИАア別 送)	運転免許証、マイナン バーカード、在留カー ド、保険証、パスポー トが読取対象。顔認証 機能あり動作に必要な ソフトウェア (JAVA) 等あり
AI OCR文字認識 カービス	東芝デジタル ソリューションズ	約4000	オムニフォント	6~40ポイント	5秒以下/1枚	Red Hat Enterprise Linux 8 64 ビット版 (x86_64)	-	システム構 成による	2021/5/1	要問合せ	
OCR Multi Entry Stage	NTTデータ NJK (メデИАアド ライブ)	手書(英数字,カタ カナ,ひらがな,漢 字4419文字,記号) 活字(英数字,カ タカナ,ひらがな, 漢字6355文字,記 号) バーコード/QR コード	手書(OCR-HN, OCR- HK, OCR-HS, OCR- B, OCR-Kなど) 活字(マルチフォ ント, IBM407, 12F)	手書(4x6mm ~12x12mm) 活字(3~ 15mm)		OCRサーバ: Windows Server 2019/2016/2 012R2/2012 クラリアン ト: Windows 11/10/8.1	-	OCRサーバ: 3GB以上 クラリアン ト: 2GB以上	2019/5/27	要問合せ	※オプシヨ ン 活字文書、名刺、運転 免許証、健康保険証

表4.2-16 サービスOCR製品（その1）

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
Cloud OCR API (領収書)	NTTデータNJK (メディアドライブ)	約3700	明朝、ゴシックなどのマルチフォント	5~45ポイント		依存なし(JSONフォーマットインターフェース)	-	2018/11	要相談	領収書OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (免許証)	NTTデータNJK (メディアドライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号(ー()・々)	明朝体、ゴシックフォント対応	42.5 ~ 236 ピクセル角程度 (500万画素カメラでカード全体がちょうど収まるサイズの 場合、1.8 ~ 10 mm 角程度)		依存なし(JSONフォーマットインターフェース)	-	2018/11	要相談	免許証OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (保険証)	NTTデータNJK (メディアドライブ)	漢字(JIS 第1水準 2965文字、第2水準 3390文字)、ひらが な、カタカナ、アル ファベット、数字、記 号	明朝体、ゴシックフォント対応	42.5 ~ 236 ピクセル角程度 (500万画素カメラでカード全体がちょうど収まるサイズの 場合、1.8 ~ 10 mm 角程度)		依存なし(JSONフォーマットインターフェース)	-	2018/11	要相談	保険証OCR Cloud商品 サーバー上での画像解析
Cloud OCR API (名刺)	NTTデータNJK (メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5~36ポイント		依存なし(JSONフォーマットインターフェース)	-	2018/11	要相談	名刺OCR Cloud商品 サーバー上での画像解析
THE 名刺管理 Business	NTTデータNJK (メディアドライブ)	約3700+英語+中韓国語	明朝、ゴシックほかマルチフォント	5~36ポイント		-	-	2018/4	800 [税別]	Cloud商品

表4.2-17 サービスOCR製品 (その2)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
クラウドOCRサービス Ver. 2.40 (AI手書き文字)	パナソニックソ リユーションテ クノロジー	<ul style="list-style-type: none"> ・英数字 ・英字、数字、記号の一部 ・住所 ・英字、数字、ひらがなの一部、カタカナ、ギリシヤ文字の一部、記号の一部、第1水準漢字の一部(2,015字)、第2水準漢字の一部(366字) ・氏名 ひらがなの一部、カタカナの一部、記号の一部、第1水準漢字の一部(2,368字)、第2水準漢字の一部(466字)、第3水準漢字の一部(8字)、JIS規格外漢字(1字) ・カタゴリフリー 英字、数字、ひらがな、カタカナ、ギリシヤ文字の一部、記号の一部、第1水準漢字の一部(2,965字)、第2水準漢字の一部(1,210字)、第3水準漢字の一部(24字)、第4水準漢字の一部(2字)、JIS規格外漢字の一部(3字) 	マルチフオント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6~60ポイント(400dpi)	-	-	-	2021/7	要問合せ	クラウドOCRサービス
クラウドOCRサービス Ver. 2.40 (名刺)	パナソニックソ リユーションテ クノロジー	<ul style="list-style-type: none"> 約6,800字 英字、数字、ひらがな、カタカナ、JIS記号(一部)168字、ギリシヤ文字(一部)32字、JIS第1水準漢字、JIS第2水準漢字 	マルチフオント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6~60ポイント(400dpi)	-	-	-	2021/7	要問合せ	クラウドOCRサービス
クラウドOCRサービス Ver. 2.40 (免許証)	パナソニックソ リユーションテ クノロジー	<ul style="list-style-type: none"> 約6,800字 英字、数字、ひらがな、カタカナ、JIS記号(一部)168字、ギリシヤ文字(一部)32字、JIS第1水準漢字、JIS第2水準漢字 	マルチフオント(明朝体、ゴシック体、教科書体、ワープロ体、新聞文字など)	6~60ポイント(400dpi)	-	-	-	2021/7	初期登録料100,000 月額利用料50,000~[税別]	クラウドOCRサービス

表4.2-18 サービスOCR製品 (その3)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
WisOCR for 注文書・請求書	パナソニックソリューションテクノロジー	手書き、活字			-	Windows 11/10	お使いのOSが推奨する環境以上	2021/12	Basic 初期登録料なし 月額基本利用料 300枚 30,000円[税別] Pro 初期登録料 100,000円 [税別] 月額 2500枚 100,000 円[税別]	PC上で動作するクラウド環境 AI-OCRエンジンで構成
RICOH 受領請求書サービス	リコー	漢字、ひらがな、英大文字、英小文字、数字、カタカナ、記号	-	-					ライトコース ベーシックコース BP0も選択可 要問合せ	
AI Read on Cloud	アライズイノベーション	活字、手書き	-	-	-	-	-	-	要問合せ	事例：注文書、請求書、決算書、アンケート用紙、チェックシート共通伝票、特許関連書類
Flax Scanner	cinnamon	対象言語：日本語、英語	-	-	-	-	-	-	要相談	事例：見積書、発注書、納品書、検収書、請求書、申込書、本人確認書類、技術文書等 提供形態：お客様環境のオンプレミス・パブリッククラウド・プライベートクラウドへの導入
CLOVA OCR Reader	LINE CLOVA	-	-	-	-	-	-	-	Template 月額 55,000円(税込) General 月額 55,000円(税込) 特化型(請求書) 月額 88,000円(税込)	事例：伝票登録OCR、領収書OCR、身分証明書OCR 詳しい利用条件は要問合せ

表4.2-19 サービスOCR製品（その4）

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
AI OCR文字認識サービス	東芝デジタルソリューションズ	約4000	オムニフオン	6~40ポイント	5秒以下/1枚	—	—	2019/1	月額利用料金30万円～ (初期構築別途)	請求書、受発注伝票等の帳票を読取対象としたクラウドサービス提供
マイナンバー収集用カード読取サービス	東芝デジタルソリューションズ	約4000	オムニフオン	6~40ポイント	4秒以下/1枚	—	—	2015/12	月額利用料金45万円～ (初期構築別途)	マイナンバーカードを読取対象としたクラウドサービス提供
本人確認書類読取サービス	東芝デジタルソリューションズ	約4000	オムニフオン	6~40ポイント	4秒以下/1枚	—	—	2015/12	月額利用料金45万円～ (初期構築別途)	運転免許証、マイナンバーカード、在留カード、保険証、パスポートを読取対象としたクラウドサービス提供
PFU Smart Capture Service	PFU Limited	活字、手書き	—	—	—	—	—	2018/4	—	アップロードされた文書画像の種類を自動的に仕分け、「特定帳票OCR」「汎用帳票OCR」「日本語手書きOCR」を自動選択して呼び出すクラウドサービス。
SmartRead	コージェントラボ	活字、手書き	—	—	—	—	—	—	スモールプラン 36万円/年 1.2万枚前後 スタンダードプラン 96万円/年 6万枚前後 エンタープライズプラン 240万円/年 26万枚前後 オンプレミスプラン	—
Tegaki	コージェントラボ	手書き	—	—	—	—	—	—	要問合せ	事例：各種申込書類やアンケートをはじめ医療機関での問診票など
スマートOCR	インフォオデオ	—	—	—	—	—	—	—	要問合せ	事例 帳票種類：決算書、健康診断書、健康保険書、勤務管理表等 オンプレミスもあり

表4.2-20 サービスOCR製品 (その5)

製品名	メーカー	認識対象文字種	認識書体	文字サイズ (400DPI)	認識速度 (字/秒)	OS	メモリ	発売年月	価格(円)	備考
DX Suite	AI Inside	活字、手書き	一	-	-	-	-		Lite [税別] 初期費用 0円 月額 30,000円～ 無料枠 18,000円分 Standard [税別] 初期費用 200,000円 月額 100,000円～ 無料枠 50,000円分 Pro [税別] 初期費用 200,000円 月額 200,000円～ 無料枠 200,000円分	Intelligent OCR Elastic Sorter Multi Form
AI よみと～ る	NTT東日本	-	一	-	-	-	-	-	プラン1 (小型) 初期費用なし 通常版 33,000円/月(税込)読 み取り箇所 月6千まで プラン2 (中型) 初期費用なし 通常版 110,000円/月(税込)読 み取り箇所 月6万まで プラン3 (大型) 初期費用なし 通常版 220,000円/月(税込)読 み取り箇所 月20万まで	プラン3 (大型) に自動帳票仕 分けオプションあり

5. ペン入力文字認識と応用

ペン入力インターフェースとは、専用のペン（又は指）とタブレット装置（又は専用の用紙）を用いて手書き筆跡をコンピューターが読み取り、文字入力やコンピューターの操作に活用する入力手法・技術である。本稿では簡単のため「ペン入力」と記すが、指によるタッチ入力も含めて考察する。ただし文字認識などを用いた情報入力手段を特に想定し、画面操作全般は含まない。

近年、スマートフォンやタブレット型コンピューターの普及により、ペン入力インターフェースは広く実用化されている。それに伴いペン入力技術そのものはコモディティー化（汎用品化）が進んでおり、新しい技術や製品の多くはペン入力デバイスの高精度化に伴うもので、既存製品の改善に留まっているように見える。本稿では、昨年度までの報告内容にいくつかの最新トピックを追加して動向の報告を行う。

5.1 ペン入力応用システム

タブレットやスマートフォンの普及が進み、筆跡を指やペンで入力するソフトウェアは製品の標準機能に組み込まれていることが多い。各社のスマートフォンの動向とペン入力技術を中心に、主な応用システムについて以下に記述する。

5.1.1 スマートフォン／タブレットの製品動向

・ GALAXY シリーズ

Samsung は専用ペン（S ペン）を使う手書き入力を前面に出した GALAXY Note を以前から発売している。その最新の後継機種が 2022 年 2 月に発表された「Galaxy S22 Ultra」である。手書き文字認識機能が標準搭載されており、また、「Galaxy Notes アプリ」によって手軽なメモ機能が実現されている。約 2 年前に発売された「Galaxy Note20 Ultra」以降、デバイス製品名に Galaxy Note という名称は使わなくなっており、「Galaxy S21/S22 Ultra」がペン入力の主力製品となっている。

Galaxy の特徴である S ペンはワコム製の電磁誘導式のペン入力機能であり、専用アプリを用いて簡単に手書きメモが取れる操作性が売りである。S ペンは新製品が出るたびに遅延が大幅に改善されている。Galaxy Note20 Ultra で 42ms から 9ms へ、Galaxy S22 Ultra では 2.8ms へ改善した[1、2]。

・ Arrows シリーズ

富士通の Arrows シリーズにはスマートフォンとタブレット端末がある。スマートフォンには独自開発の手書き入力機能が搭載されている。ATOK ソフトキーボード上で筆跡入力可能なシームレス入力に加え、一つの文字枠内に文字を重ね書きできる機能も搭載されている。

2021年モデルの Arrows We (F-51B) 等にも踏襲されている[3、4]。タブレットの最新機種は Arrows Tab F-02K である。2018年のモデルだがそれ以降の更新は行われていない[5]。

- HUAWEI タブレット

ファーウェイ・ジャパンは2017年8月に筆圧感知に対応した8インチタブレット「MediaPad M3」を発売し、その後「MediaPad T5」（2019年8月）や「MatePad」（2020年6月）など、様々な製品を発売している。付属の専用ペン「M-Pen」による筆記が可能で、後述の MyScript Nebo が標準搭載されている。シリーズ中の最新機種は2021年7月に発売した「MatePad 11」であり、HUAWEI M-Pencil（第2世代）を採用。圧力・傾斜を4096レベルで感知し、約2ミリ秒の遅延を実現している[6、7]。

更に2022年6月に、E-Ink社の電子ペーパーを採用した「HUAWEI MetaPad Paper」が発売された。独自の「HarmonyOS 2」を搭載し、独自の電子書籍ストアアプリ「HUAWEI Books」も導入された電子書籍端末だが、富士通の「QUADERNO」等と同様にスタイラスペンで手書きが可能なのが特徴である。電子書籍への手書き機能は充実していないとの報告があるが、手書きメモ機能としては十分に使い、高精度な手書き文字認識機能も搭載されているようである。なお、ペン入力機能は第2世代の M-Pencil を採用している[8]。

5.1.2 スマートフォン／タブレット向け手書きソフトウェア

- OneNote[9、10、11]

OneNoteは、マイクロソフト社が提供する無料の Windows/Mac/iOS/Android/Web アプリケーション向けのデジタルノートアプリケーションである。オフライン手書き文字認識機能があり、ペン入力した手書き文字や、カメラで撮影した手書き文字などを認識する。OneNote 各OS用のアプリストアから無料でダウンロードできる。Windows版は、従来は32bit版と64bit版が別々に提供されていたが、2022年10月から32bit版が「Windows版」として統一提供される。これはWindows11に同梱されるものと同じであり、また、Microsoft 365を導入した場合も自動でセットアップされる。

- Neo Studio[12]

Android、iOSのスマートデバイス用アプリケーション Noe Studio は専用の微細なコードが印刷された紙のノートと専用のネオスマートペン（M1、N2、Dimoの3タイプ）を使うことで、手書きのリプレイやテキストへの変換などができるアプリである。テキスト変換は、英語、日本語、中国語等15か国の言語をサポートしている。

- MyScript Nebo[13、14]

MyScript Neboは、iOS、Windows、Android上で使えるノートアプリである。一時期有料となったが、現在は無料となっている。手書き変換機能があり、リアルタイムに手書き文字が

文字認識されていく。書いた後に文字をダブルタップするとテキストに変換する機能や、PDF ファイルをインポートして注釈を付ける機能もある。2022 年 12 月に最新の Nebo 4.0 がリリースされ、手書きジェスチャ等が使えるようになった。

- mazec[15]

サードパーティー製のソフトキーボードには、GALAXY に採用されている「mazec」がある。Windows、Android、iOS に対応しており、個人向けアプリに加えて法人向けの mazec for Business や業種向け（医療、建設等）、SDK などもリリースされている。最新辞書を同期できるクラウドサービス mazec Plus が iOS 向けに存在する。

5.1.3 クラウドサービス

- Google 入力ツール「手書き入力」[16、17、18]

Google のテキスト入力機能を WebAPI で利用できるようにしたクラウドサービスである。仮名漢字変換や仮想キーボードなどの文字入力機能を持ち、その中に「手書き入力」も含まれる。手書き入力機能は、マウスやタッチの移動情報を時系列の座標に変換し、それを json 形式にしてサーバーに送ると文字候補一覧が json 形式で返ってくるという仕組みである。現時点で 118 言語に対応している（2023 年 1 月現在：2020 年 1 月から変化なし）。

5.1.4 ペン入力デバイスの動向

スマートフォンやタブレット筆記デバイスは指とペンの併用が可能な静電誘導式タブレットが使われることが多いが、筆跡データ入力を重視した用途が広がるにつれ、専用ペンを搭載するケースも次第に増えている。

近年の製品に使われている専用ペンは主に 4 種類の方式（プロトコル）が使われている[19]。Wacom には EMR 方式と AES 方式があり、Microsoft は MPP (Microsoft Pen Protocol)、Apple も独自方式を採用している。ASUS 等が採用していた Synaptic 方式は、最近はあまり見ないようである。

各プロトコルの概要を下記に記す。どの方式が良いかについてはいくつかのレビュー記事が存在するが、概ねプロ用（イラスト等）の書き味としては Wacom EMR と Apple Pencil を推す声が多く、筆圧や傾き検出などの機能では Wacom EMR と MPP が優っているという声が多い。

- Wacom EMR 方式[20]

ワコムが特許を取得している電磁誘導方式（EMR: Electro-Magnetic Resonance）であり、同社の MobileStudio Pro 等で使われているプロ仕様の規格である。Wacom Feel IT Technologies という名称で呼ばれることもあるが、これは後述の Wacom AES 方式でも使われ

ることがあり、呼称や規格名は必ずしも統一されていない。EMR 方式の中にも 4 種類の規格があるなど、互換性については個別に確認する必要がある[21]。

- Wacom AES 方式[22、23]

ワコムのアクティブ静電結合方式 (AES: Active Electrostatic) である。他社 PC に採用されているのは多くがこの方式であり、公式の Wacom 「Bamboo Ink」だけでなく、互換性のあるペンが他社から発売されている。

- Microsoft MPP 方式

マイクロソフトの Surface Pro 3 で採用された「Surface Pen」が用いている方式である。マイクロソフトは Surface Pro 2 まではワコムの AES 方式を用いていたが、イスラエルの N-Trig 社を買収し、その技術を導入した方式を MPP (Microsoft Pen Protocol) と称している。

- Apple 方式

Apple Pencil が採用しているアップル社独自の方式である。

続いて、専用ペンの主な製品を記す。Wacom EMR や Apple Pen はメーカーが固定されるため上記の方式と分類が重なる部分が多い。両方の説明を併せて参考にして欲しい。

- Surface Pen

Microsoft Surface Pro 3 で採用された「Surface Pen」は、前身の Surface Pro 2 で「Pro Pen」と呼ばれていた専用ペン (Wacom 製) のバージョンアップ版であり、電池を用い Bluetooth 機能を内蔵している (イスラエルの N-Trig 社製) [24、25]。2017 年 8 月発売の「Surface Pro」(Surface Pro 4 の次機種なので「Pro 5」と呼ばれる) は、筆圧検知が 1024 段階から 4096 段階に拡張され、傾き検知にも対応した新「Surface Pen」が採用された[26]。Surface Pen は PC 本体とは別売りであり、同じ PC で旧ペンも新ペンも使える[27]。

Surface Pen は使用する通信プロトコル MPP (Microsoft Pen Protocol) のバージョンが PC と一致している必要がある。例えば、最新機種である Surface Pro 9 は「Surface スリム ペン 2 専用」と仕様に記されている。スリムペン 2 は MMP v2.6 であり、4,096 段階の筆圧検知、傾き検知などの機能をサポート対応している[28]。MPP のペンはいわゆる静電容量方式で、ペン側で電磁波を発生させることで、パネル側がペンの位置を検知する。このためペンの側には Apple Pencil と同じようにバッテリーが必要になるが、ペン先に電磁波を発生させるだけでいいので、小さな電池 (単 6 形) で半年といった長期間利用ができる[29]。

- Apple Pencil

Apple Pencil は iPad Pro の入力デバイスとして 2015 年 11 月に発売され、2018 年 10 月に第二世代が発表された。2018 年以降の iPad や iPad mini では第一世代 Apple Pencil が使える

が、第二世代 Apple Pencil は iPad Pro のみ対応である。第二世代の Apple Pencil は形状や充電方法などの使い勝手が第一世代より改善されている[30]。[30]は 2021 年の記事だが、2023 年 1 月現在、Apple Pencil 第三世代は発表・発売されていない。しかし細かな仕様について言及された記事が目立っており、2024 年秋の iPad Pro 新版とともに第三世代が発売されるのでは？と噂されている[31]。

Apple Pencil にはサードパーティー（ロジクール）製の Crayon というペンもあり、Apple Store では 3 種類のデジタルペンが販売されている[32]。なお、Apple Pen は iPad Pro 本体とのデータのやりとりに Bluetooth LE（Low Energy）を利用するため、ペンのバッテリーは Surface Pen ほど持たない。（Apple Pencil は 1 回の充電で約 12 時間、Crayon は約 7.5 時間）

- Wacom Pro Pen 2

ワコムは、2016 年 11 月に発売した「Wacom MobileStudio Pro」において、筆圧 8192 段階の「Wacom Pro Pen2」を採用した[33]。これは 2017 年 1 月発売のタブレット「Intuos Pro」を始め、多くのワコムタブレットに採用されている。2019 年 2 月には同じ性能でスリムタイプの「Wacom Pro Pen slim」も発売されている[34、35]。

- GALAXY Note 「S ペン」 [1, 2]

GALAXY Note には「S ペン」という電磁誘導式の専用ペンが添付されている。これはワコムの技術を採用しており、筆圧も感知できるため自然な筆跡の入力（太さの変化まで表現）が可能である。S ペンは Galaxy シリーズの強力な特長となっており、新機種が発表されるごとに改善が加えられている。2022 年発売の Galaxy S22 Ultra では S ペンは遅延時間を約 2.8 ミリ秒まで短縮し、ペンを本体に格納できるようになった [36]（5.1.1 節参照）。

5.1.5 その他の製品・サービス

- SkyCom 社 SkyPDF Touch Ink for win 7 [37]

同社製品「SkyPDF Professional 7」（PDF 作成・編集・加工を行う）のオプション製品であり、Windows タブレット上に表示した PDF に直接ペンで手書き文字入力するアプリケーションである。紙に文字を書くような自然な書き味を再現し、手書きした文字は手書き文字認識機能によりテキストデータに変換し、用途に合わせて置換や埋込が簡単に行える。また、手書き筆跡に電子署名技術を融合させ、証跡情報を暗号化して埋め込むことで、ペーパーレスを推進するソリューションも提供している。

5.2 教育向け応用

- コンシューマ向け製品

オンライン手書き認識を活用した電子教材は、かつてタブレット PC の普及に伴っていくつ

かのシステムが開発された。例えば、富士通研究所では独自開発の筆順画数フリー手書き漢字認識・評価エンジンを用いて 2003 年度に手書き電子教材の実践研究を行い、その成果が小学館の「デジタルドラゼミ」(2013 年 7 月 31 日終了)や「小学館 徹底反復デジタルドリル」(2019 年 3 月終了)に採用された。この電子教材システムは、富士通の文教ソリューションの一環として提供中である[38]。また、ニンテンドーDS 上で「漢検 DS」や「美文字トレーニング」等が発売された[39、40]。それぞれ 2006 年、2008 年発売だが、新しいニンテンドー3DS でも動作する。

5.3 主な日本語オンライン文字認識エンジン

- ・ iLabo 手書き文字認識エンジン[41、42]

東京農工大学中川研究室が開発したオンライン手書き文字認識技術を事業化するために設立された大学発ベンチャー「アイラボ株式会社」が販売している認識エンジンである。中川研究室のオンラインマッチング（高速な非線形伸縮マッチング）とオフライン文字認識（OCR）を統合したハイブリッド型文字認識を用い、CRF 等を用いた学習技術も含む。MetaMoji 社の mazel や、Samsung GALAXY シリーズ、7note 等に採用されている。

- ・ 富士通オンライン文字認識エンジン

富士通製スマートフォン（Arrows シリーズ）に搭載。手書き電子教材でも利用している。オンラインマッチングとオフライン文字認識を統合したハイブリッド型文字認識である。オフライン認識は富士通製 OCR の技術を用いており、独自の文脈処理技術や予測機能も有する。

- ・ MyScript[43]

多言語に対応したオンライン文字認識エンジンを展開しているフランスの会社である。従来は MyScript 社が Vision Object というブランドでソフトを開発していたが、現在は会社名もブランドもすべて MyScript に統一されている。

- ・ Ink シリーズ（ポトス株式会社）

ポトス株式会社は 1997 年に設立された、ペン入力インターフェース技術を中心にライブラリやソリューション開発を行っている企業である[44]。InkTool（ペン入力ソフト開発ツール）や InkFep（日本語手書き文字認識ツール）などが様々な企業で利用されている。例えば、富士通は自社ソリューションにおいて手書き文字入力機能の実現をサポートする開発ライブラリとして InkTool 及び InkFep を活用している[45]。ポトス株式会社の製品は、文字認識機能そのものは東京農工大学（中川研究室）の技術を利用している。

- [20] ワコムのテクノロジー Electro-magnetic Resonance
<https://www.wacom.com/ja-jp/for-business/technologies/emr>
- [21] ワコムの電磁誘導ペン（EMR）は現在 4 種類もある上に全部互換性がなくてめちゃ分かりづらいのでまとめておく
https://29udon.com/6143.html#Wacom_Feel_IT_TechnologiesEMR
- [22] ワコムのテクノロジー Active Electrostatic
<https://www.wacom.com/ja-jp/for-business/technologies/aes>
- [23] Wacom AES 対応デジタルペンの“最強”はどれだ？
<https://mupon.net/wacom-aes-no1-pen/>
- [24] Microsoft、マルチタッチスクリーン技術の N-trig に出資
<http://www.itmedia.co.jp/news/articles/0901/13/news029.html>
- [25] Surface Pro 3 と Pro 2 のスタイラスペン比較（N-Trig vs Wacom）
<http://tabkul.com/?p=64938>
- [26] 新しい「Surface Pro」は「4」よりも完成度が高まった！
<https://kagakumag.com/pc-smartphone/?id=10536>
- [27] [新 surface pen 描き心地レビュー] 絵・イラストの描き味は旧ペンとどう変わった？
<https://maekoart.net/new-surface-pen>
- [28] 別の PC で Surface ペンを Windows する
<https://support.microsoft.com/ja-jp/surface/7e1861d0-d6fa-4ba5-a9e3-fe210806b211>
- [29] 同じ 10 型級の Surface Go と iPad Pro はどちらが使いやすいのか？多方面から実機で検証
<https://pc.watch.impress.co.jp/docs/column/ubiq/1140203.html>
- [30] [比較] Apple Pencil の第 1 世代・第 2 世代のどちらを買えばいいの？
<https://yossense.com/comparing-apple-pencils/>
- [31] 【最新情報まとめ】Apple Pencil 第 3 世代 発売日・スペック
<https://motifyublog.com/new-apple-pencil-spec/>
- [32] Apple Pencil とロジケールの Crayon、どちらを買うか本気で考えた
<https://news.mynavi.jp/article/20190617-843972/>
- [33] ワコム、思い通りの制作フローを可能にする Wacom Intuos Pro を発売
<http://www.wacom.com/ja-jp/about-wacom/news-and-events/2017/1213>
- [34] 「ペンが走ってお絵描きが楽しい」ワコムの液タブ・ペンタブ用のスリムペン「Wacom Pro Pen slim」レビュー <https://gigazine.net/news/20190227-wacom-cintiq-pro-pen-slim/>
- [35] くらべてみました！ Wacom Pro Pen
<https://tablet.wacom.co.jp/article/choice-wacompropen>

- [36] Samsung が「Galaxy S22 シリーズ」を発表 Ultra は「S ペン」を収納可能に
<https://www.itmedia.co.jp/mobile/articles/2202/10/news052.html>
- [37] Windows タブレットを使って PDF に手書き文字入力 / 電子サイン「SkyPDF Touch Ink for win 7」
https://www.skycom.jp/product/skypdf/touchink_for_win_7/
- [38] FUJITSU 文教ソリューション K-12 個別学習支援 ペンまる V1
<https://www.fujitsu.com/jp/solutions/industry/education/school/learning/penmaru/>
- [39] 手書きの文字が持つ大きな魅力 <http://www.nintendo.co.jp/nom/0803/p1/>
- [40] 財団法人 日本漢字能力検定協会公認 漢検 DS
<http://www.nintendo.co.jp/ds/software/akjj/>
- [41] 日本語文字列認識エンジン
https://ilabo.biz/ilabo_japanese_engine/
- [42] 世界最高精度の手書き文字認識技術を実用化!
https://www.jst.go.jp/pr/jst-news/backnumber/2012/201205/pdf/2012_05_p12.pdf
- [43] MyScript Labs 手書き認識のリーダー
<https://www.myscript.com/ja/handwriting-recognition>
- [44] ポトス株式会社 <http://pothos.to/>
- [45] セキュリティ Solution ラインナップ (その他)
<https://www.fujitsu.com/jp/solutions/business-technology/security/secure/lineup/solutions-2/> (URL 確認 2023.3.23)

6. 海外動向調査

6.1 文書画像認識に関する主な国際会議

文書画像認識の研究分野は、IAPR (International Association of Pattern Recognition : 国際パターン認識連盟) が国際的なコミュニティの中心となっている。その中でも、TC 10 (Technical Committee Number 10 : 第 10 技術委員会 : “Graphics Recognition”) [1]と、TC 11 (Technical Committee Number 11 : 第 11 技術委員会 : “Reading Systems”) [2]が特に文書画像認識に関係する。IAPR TC10/11 が主催する主な国際会議には ICDAR、DAS、ICFHR があり、これらが文書画像認識の国際的な研究動向を把握するために特に重要である。それぞれの国際会議の特徴を下記に示す。

●ICDAR (International Conference on Document Analysis and Recognition)

文書認識・解析の全般に関する最大の国際会議である。1991 年から隔年開催で、2021 年に第 16 回がスイスのローザンヌで開催された (オンラインと現地のハイブリッド開催) [3]。近年の参加者は 500 名前後である。ICDAR2021 における投票により [4]、ICDAR は次回から毎年開催することとなった。次回 2023 年の第 17 回 (米国・サンノゼ) [5]の次は 2024 年 (ギリシア・アテネ) で開催される。

●DAS (International Workshop on Document Analysis Systems)

ICDAR に比べて規模は小さいが、扱う範囲はほぼ同じである。1994 年から隔年開催で、今年度は 2022 年 5 月にフランスのラ・ロシェルで開催された (ハイブリッド開催) [6]。Workshop と銘打たれており、シングルトラックですべての口頭発表を聞くことができる。参加者は 150 名程度。

●ICFHR (International Conference on Frontiers in Handwriting Recognition)

手書き文字認識に関する国際会議である。1990 年に Workshop として (名称は IWFHR) 第一回が開催され、ほぼ隔年で実施されている。規模の拡大に伴い、2008 年の第 11 回から名称が Conference となり、ICFHR に改称。2022 年 12 月にインドのハイデラバードで開催された (ハイブリッド開催) [7]。参加者は DAS とほぼ同じ 150~200 名程度。

6.2 国際会議の再編成についての議論

これまでは奇数年に ICDAR、偶数年に DAS が行われてきたため、文書画像認識全般の技術動向を追跡するためには ICDAR 及び DAS へ交互に参加していれば良かった。しかしながら、2023 年以降は ICDAR が毎年開催されるため、ICDAR のみに注目すれば従来と同様の動向把握は可能である。それに伴い、ICDAR とそれ以外の関連する会議の立ち位置を見直す必要があるのではな

いか、との議論も行われている[8]。

上記の議論は ICDAR 中の“Future of DAS Workshop”で行われ、現在は“Kialo”という議論用サイトでオープンな議論が継続している。議論のテーマは“ We should fuse ICDAR / ICFHR / DAS / GREC into a single, annual conference.”（我々は ICDAR / ICFHR / DAS / GREC を単一の年次開催の会議に融合すべき）であり、Pros and Cons（賛否）のコメントが数多く寄せられている。その延長上で、IAPR TC11 のメーリングリストでは、2022 年 12 月のメールにおいて、ICFHR を単独開催とするか ICDAR に統合するかの投票が行われた[9]。主な選択肢は下記のとおりである。

- 各 Workshop は従来と同様に単独で開催する（毎年 or 隔年）
- ICDAR と統合して、ICDAR の Satellite Workshop として開催する

ICDAR と統合の統合に賛成の意見には、例えば、DAS は System only ではなく ICFHR は Hand-Writing only ではなく、ICDAR との差異は曖昧なので統合すべきというものや、より大きな会議として世の中の注目を集めたいというものがある。一方で ICDAR がより大きくなれば運営は困難となり開催地への立候補のハードルが上がると危惧する意見や、ICFHR や GREC にはより狭く深い技術にフォーカスした発表があるので、個別に実施した方が有意義だという意見もある。議論の流れや方針は逐一変わるので、TC10/TC11 の Newsletter を把握しておくが良い。

6.3 2022 年度の開催状況

2022 年度は、5 月に DAS2022 が、12 月に ICFHR2022 が開催された。2020 年度から拡大した新型コロナウイルス感染症（COVID-19）の世界的な流行のため、一時期にはすべての会議がオンラインのみとなっていたが、今年度には現地とオンラインのハイブリッド開催が標準となってきた。上記の 2 会議もハイブリッド開催であった（参考：DAS2020 と ICFHR2020 はいずれもオンライン開催）。

6.3.1 国際会議 DAS2022

表 6.3-1 DAS2022 の開催概要

日程	2022年5月22日(日)～5月25日(水)
Tutorial	5月22日(日)
Main Conference	5月23日(月)～5月25日(水)
開催地	フランス ラ・ロシェル
会場	ハイブリッド開催(現地会場は下記) La Rochelle University
主催	IAPR (International Association of Pattern Recognition)
公式サイト	https://iapr.org/archives/das2022/index.html

表 6.3-2 前回大会との比較

	前回 (DAS2020)	今回 (DAS2022)
開催期間	4日間	4日間
開催時間/日	約6時間半	約9時間 (Lunch Break 含む)
Tutorial	1件(約3時間)	1件(約2時間半)
Keynote Speech	3件	2件
Poster Session	1枠	2枠
Oral Presentation	6枠	6枠

前回大会との比較を表に示した。前回はコロナ禍でのオンライン開催のため開催時間が短めに変更されていたが、今回は現地時刻で9:00前後から18:00前後まで、コロナ前と変わらない時間帯で開催された。しかし、最終日の終了時刻が早い(15:00終了)、Oral Session数が少ないなど、ハイブリッド開催を考慮した調整も行われたようである。しかし、DASの特徴であるディスカッショングループは行われた。今回、本稿の報告者は会議に参加していないため様子は分からないが、昨年の報告書にも記したようにハイブリッド開催では自由な議論は難しいので、今回も運営は難しかったであろう事は想像できる。

[Tutorial]

- Tutorial (Himanshu Sharad Bhatt)

“Unlocking the Potential of Unstructured Data in Finance Through Document Intelligence”

[招待講演 (Keynote Speech)]

- Keynote 1 (Prof. Adam Jatowt)
“Automatic Question Answering & Generation in News Archives”
- Keynote 2 (Prof. Andreas Dengel)
“Bringing Document Content to Life”

[口頭発表]

- Oral Session 1: Document Analysis Systems and applications (5 件)
- Oral Session 2: Information extraction and applications (4 件)
- Oral Session 3: Historical document analysis (7 件)
- Oral Session 4: Handwriting Text Recognition (4 件)
- Oral Session 5: Applications in handwriting (6 件)
- Oral Session 6: Open-source software and Benchmarking (5 件)

[その他]

- Industrial track
- Discussion group

[ポスター発表]

- Poster: Session 1 (Full paper 10 件、Short paper 7 件)
- Poster: Session 2 (Full paper 10 件、Short paper 8 件)

[参加報告]

オンサイトとオンラインのハイブリッドで開催された。ポスターセッションもオンサイト／オンラインの交代制で実施され、オンラインでのセッションは **gather.town** 上での開催であった。

ここ数年で開催者にオンラインセッション開催のノウハウが蓄積されてきていることが感じられた。前回 DAS2020 はオーラル、ポスターとも **Zoom** での開催だったが、今回はメタバース上での開催となっており、進化が感じられた。特にポスターセッションでは現在の集客状況が事前に分かりやすく、特定のポスターに人が集中することを避ける効果があったように思われる。一方で現地参加者も増え教室が埋まるくらいになっており、国外でもコロナが落ちついてそれ以前に回帰しつつあることが見て取れた。



図 6.3-1 オンラインポスターセッションの画面

発表内容には文書からの情報抽出（Question Answering : QA 含む）や構造化に関するものが多く見られ、Keynote でも University of Innsbruck の Dr. Adam Jatowt（元京都大学）による QA に関する講演（Automatic Question Answering & Generation in News Archives）があった。研究の中心は明らかに文書データの活用に移行しつつあり、文字認識を直接的には経ない方法が主流となり始めている。すでにマルチリンガルな QA も扱われており、このような技術を活用したサービスの登場も想定以上に早い可能性がある。今後もこの分野の注視が必要である。

また、歴史文書の発表もコンペティション関連を含めて多かった。この分野の各コンペティションタスクでマルチタスク学習の検証を行った例などもあり、活況を呈していた。歴史文書の認識は研究での活用を中心に国内でも一定のニーズがあり、今後も継続していくものと思われる。

6.3.2 国際会議 ICFHR2022

表 6.3-3 ICFHR2022 の開催概要

日程	2022 年 12 月 4 日（日）～12 月 7 日（水）
Workshop	12 月 4 日（日）
Main Conference	12 月 5 日（月）～12 月 7 日（水）
開催地	インド ハイデラバード
会場	ハイブリッド開催（現地会場は下記） The International Institute of Information Technology, Hyderabad
主催	IAPR（International Association of Pattern Recognition）
公式サイト	http://icfhr2022.org/

表 6.3-4 前回大会との比較

	前回 (ICFHR2020)	今回 (ICFHR2022)
開催期間	4 日間	4 日間
開催時間／日	約 3 時間 40 分	約 9 時間 (Lunch Break 含む)
Tutorial (or Workshop)	1 件 (約 3 時間)	2 件 (約 7 時間半)
Keynote Speech	3 件	3 件
Poster Session	0 枠 (並行開催)	1 枠 (Panel Discussion)
Oral Presentation	6 枠	8 枠

ICFHR についても前回との比較を表に示した。前回はフルオンラインで会議時間も発表数も大幅に縮小されたが、今回は元に戻っている。ポスターセッションが無く、代わりにパネルディスカッションが追加されるなど、内容の構成は少し変化しているように見える。セッションが技術テーマで分けるのではなく、Online/Offline/Mixed のように参加形態で分けられているのが特徴的である。

[Workshop]

- Workshop on Improving Document Analysis Tools for Indian Languages
- 1st Workshop on Deep Document Understanding

[招待講演 (Keynote Speech)]

- Keynote 1 (Dr. Christopher Kermorvant)
“Large-scale handwriting recognition: implementation in historical document processing projects”
- Keynote 2 (Dr. Prem Natarajan)
“Frontiers of Fair and Accessible (Conversational) AI”
- Keynote 3 (Prof.Rajeev Sangal)
“Document Understanding for Machine Translation”

[口頭発表]

- Oral Session 1: Conference Track -Offline (8 件)
- Oral Session 2: Journal Track -Offline (4 件)
- Oral Session 3: Conference Track -Mixed (8 件)
- Oral Session 4: Conference Track -Online (8 件)

- Oral Session 5: Conference Track -Online (5 件)
- Oral Session 6: Journal Track -Online (5 件)
- Oral Session 7: Presentation of papers published elsewhere -Online (4 件)
- Oral Session 8: Conference Track -Online (8 件)

[その他]

- Panel Discussion: “Intelligent Technologies And Systems For Palm Leaf Manuscript Understanding”
- Remembering Prof Sri Hari

[参加報告]

オンサイトとオンラインのハイブリッドで開催された。投稿・採択数が少なかったためか全件オーラルでの発表となり、ポスターセッションはなし。ICDAR2021 などと比べて企業の参加が少なく、また、開催地の関係から Urdu 語に関する発表が比較的多く見られた。

QA に関する発表もあったが、会議テーマの上からも手書き文字認識の発表が多数を占め、特に Vision Transformer (ViT) を応用した研究が流行していた。ベストペーパーは CTC 系の手法が獲得していたが、全体的には ViT への転換が進んでいく可能性がある。

アプリケーションよりの話として、東京農工大の中川研究室から、記述テストの自動採点に関する発表があった。共通テスト導入に向けたプレテストの答案データセット (NCUEE-HJA) を活用した研究であり、採点は自然言語処理 (NLP) 技術により行っている。現状ではまだ採点精度に課題はあるが、今後日本でも学習・試験への OCR・NLP 技術導入が進んでいくと思われる。

【参考文献】

- [1] IAPR TC10 Homepage <https://iapr-tc10.univ-lr.fr>
- [2] IAPR TC11 Homepage <http://www.iapr-tc11.org>
- [3] ICDAR 2021 Homepage <https://icdar2021.org>
- [4] [IAPR-TC10] Newsletter 147 – September 2021 <https://iapr-tc10.univ-lr.fr/?p=1366>
- [5] ICDAR 2023 Homepage <https://icdar2023.org>
- [6] DAS 2022 Homepage(Archive) <https://iapr.org/archives/das2022/>
- [7] ICFHR 2022 Homepage <http://icfhr2022.org/>
- [8] We should fuse ICDAR / ICFHR / DAS / GREC into a single, annual conference.
<https://www.kialo.com/we-should-fuse-icdar--icfhr--das--grec-into-a-single-annual-conference-30656>

7. 今後の展望

認識形入力方式は、我が国で急務となっているデジタル化を推進するための重要な基盤技術であり、今後も急速な勢いで進化し続けるものと予想される。一方、技術進化に伴いその活躍の場が拡大することで、利用環境が多様化し、環境に起因する様々な外乱の影響を受けやすくなっている。外乱が認識形入力方式を搭載したシステムの性能を低下させる可能性があることは知られているものの、これまで明確な整理は行われてこなかった。これらの外乱要因とその影響を合理的に整理することができれば、認識形入力方式を搭載した機器の技術者やユーザーに気づきを与え、その利活用を促進できるだろう。要因表及びその活用方法については引き続き議論を進め、ガイドライン・報告書などでの公開に向けて、ターゲットとなる読者に合わせた表現でまとめる作業を行う。

既に実用化が進んでいる整備された環境におけるOCR装置については、引き続き最新情報のアップデートを行い、我が国におけるこの分野の信頼できる情報発信拠点として公開を行っていく。また、関連する規格について各規格の利用状況を考慮しつつ、見直し等を進める。

DL等の技術革新を背景とした認識形入力方式の最新技術動向や、これを搭載したシステム及びサービスについても動向調査を継続するとともに、最新技術の実用化に向けた課題、及び急速に進化する技術と拡大する市場の要求レベルを調和させる各種規格化について議論を深めていく予定である。

— 禁無断転載 —

認識形入力方式に関する調査研究報告書

発行月	2023年3月
編集・発行	一般社団法人 電子情報技術産業協会 認識形入力方式標準化専門委員会 〒100-0004 東京都千代田区大手町 1-1-3 大手センタービル TEL (03) 5218-1058