

認識形入力方式に関する調査研究報告書

2022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、デジタル技術とデータを活用して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。政府もデジタル庁を設置するなどその推進体制を強化しているところであるが、これに期せずして新型コロナウイルス感染症の影響が拍車をかける形となり書類の電子化、押印廃止、テレワーク、オンライン会議など業務のデジタル化が急速に進み、普及することとなった。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の入力として用いる仕組みは、実世界の様々な情報をデジタル世界に汲み上げるための極めて重要な基盤技術であるといえ、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考えられる。

このような背景のもと、当協会では、2021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を、現象・原因の観点から網羅的な整理を行うとともにその構造化を進め、開発者やユーザーにとって有用なガイドラインとしていくための議論を行った。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。更に、今年度開催された学術会議等の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。

2022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 佐藤 雄 隆

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	佐藤雄隆	国立研究開発法人 産業技術総合研究所
副委員長	高橋勝彦	日本電気株式会社
監事	山合敏文	株式会社 リコー
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	田中宏	富士通株式会社
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

(2022年3月31日現在)

目 次

1. はじめに	1
1.1 審議過程	1
1.2 技術の現状調査	2
2. 実世界環境における認識機器の耐環境性の標準化	3
2.1 実世界環境における外乱要因	4
2.2 デジタル化に関する俯瞰的考察	9
2.3 RPAにおけるUI操作自動化のためのOCRの外乱要因	12
2.4 実世界環境における認識機器の現状と今後の展望	17
3. 認識技術の動向	27
3.1 文字認識技術の現状と今後の展望	27
3.2 文字認識・文書理解に関する国内学会の発表動向	32
4. 文字認識システムの技術調査	39
4.1 OCRの現状	39
4.2 製品分類について	43
5. ペン入力文字認識と応用	66
5.1 ペン入力応用システム	66
5.2 教育向け応用	71
5.3 主な日本語オンライン文字認識エンジン	72
6. 海外動向調査	77
6.1 文書画像認識に関する主な国際会議	77
6.2 2021年度の開催状況	78
6.3 国際会議 ICDAR2021	79
7. 今後の展望	83

1. はじめに

いわゆる「デジタル化」は、政府の従来からの重要政策である「働き方改革」の文脈においても重要課題と位置付けられ、2021年9月にはデジタル庁が設置され推進体制が強化された。これに、新型コロナウイルス感染症の影響も加わり、デジタル化の推進は正に待ったなしの状況となっている。認識形入力方式は実世界の情報をデジタル世界に汲み上げるための重要な基盤技術であり、デジタル化においても非常に重要な役割を担うことになる。一方、様々な実環境にその応用の範囲が拡大するにつれて、多様な外乱の影響を受けることになり、条件によっては性能の劣化やユーザビリティの低下が起こることがある。このため本委員会では、実環境における外乱要因を現象・原因の観点から網羅的に整理するための議論を行った。また、急速に進化するAI技術を背景として発展が著しい認識方式と認識技術の動向調査についても製品や技術の観点から行った。

委員会は2021年5月から2022年2月まで計8回オンライン開催され、下記のような審議が行われた。

1.1 審議過程

1.1.1 実世界環境における認識機器の耐環境性の標準化

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年度本に委員会からリリースしたが、引き続きOCRに影響を及ぼす外乱を現象・原因の観点から網羅的に洗い出す作業を進めた。認識形入力方式は実環境の様々な場面で用いられるため、多様な外乱の影響を受け場合によっては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。一方、もし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。このような狙いから、本年度当委員会では、実環境における外乱要因を現象・原因の観点から網羅的な整理を行うとともに、その構造化を進め、開発者やユーザーにとって有用なガイドラインとしていくための議論を行った。また、「デジタル化」というキーワードが俯瞰的視点から見た場合に、具体的に何を意味するのか、OCRやRPAの文脈を中心としながら議論・整理を行い、あくまで議論の途中過程ではあるが、本報告書中にその整理の案を掲載した。

1.2 技術の現状調査

1.2.1 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、オムロンサイニックエックス社の牛久 祥孝氏に「Transformer」、産業技術総合研究所の大岩 寛氏に「機械学習品質マネジメントガイドライン策定と標準化の取組み」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いただいた。

1.2.2 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。

1.2.3 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

我々の身の回りにある情報の所在は、①実世界と②サイバー世界に大別することができる。コンピューターは、サイバー世界に存在する情報を誤りなく高速に処理可能であるが、実世界に存在する情報を同様に処理するためには、何らかの手立てによってサイバー世界の情報に変換する必要がある。その最も単純な実現方法は人手に頼ること（例えば、文書・帳票に記載された文字情報をキーボードで入力するなど）であるが、この方法は人的コストが高いという問題がある。そこで、登場したのが認識機器である。コンピューターの登場以来、認識機器も多くの場合コンピューターを用いて実現されるが、バーコードなどの機械処理用に特別に設計された情報を除いて、実は実世界の情報をサイバー世界の情報へ誤りなく変換する完璧なアルゴリズムは未だ発見されていない。そこで、認識対象、すなわち認識機器が変換対象とする実世界の情報を認識容易なものに制限したり、認識誤りの原因となる外乱要因が途中過程に入り込まないように装置を工夫したりすることで実用に供する性能を実現している。

OCR は実世界に存在する文書・帳票中の情報（文字）を認識し、サイバー世界の文字コードに変換する認識機器である。従来の OCR は、認識対象とする文字の字形や印字・筆記の条件、さらには OCR 読み取りのための専用帳票を規格で定め制限していた。また、実世界からサイバー世界へ情報変換する前半の過程であるデジタル画像化の過程においては、光学的及び幾何的な外乱要因が入り込まないスキャナー機構を用いることで高精度な文字認識を実現していた。しかし近年では、認識対象拡大に対するニーズ、DL に代表される認識技術の進歩（シーズ）、スマートフォンの普及などが相まって、スマートフォン内蔵のカメラで撮影されたレシートや請求書なども認識できる OCR が登場している。また、ホワイトカラーの労働生産性向上に対する強い要請から、UI 操作を自動化するための手段として PC 画面に表示された文字を認識する用途にも OCR が活用されるようになりつつある。

このように OCR の活用範囲が拡大していることは喜ばしいことであるが、その一方で注意すべきことがある。上述のように、実世界の情報をサイバー世界の情報へ誤りなく変換する完璧なアルゴリズムは未発見なため、認識技術が進歩したとはいえ、OCR を適切な環境下で利用しないと期待通りの認識が行えない可能性がある。ところが、これまで認識精度に影響を与える具体的な外乱要因についてはほとんど明らかにされていなかった。この状況を放置すると OCR が適切な環境で利用されない事態を生み出しかねない。

当委員会ではこのような状況に鑑み、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平成 24 年度より議論を開始した。まず、実世界環境での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て、OCR の利用ガイドラインを昨年度 4 月に制定した。また、実世界環境下で認識精度に影響を及ぼす様々な外乱要因及びそれにより生じる画像劣化の現象に

について網羅的な列挙を開始し、今年度はそれらについて一通りの整理を行った。並行して RPA における UI 操作自動化のための OCR に影響を及ぼす外乱要因について議論を進めた。

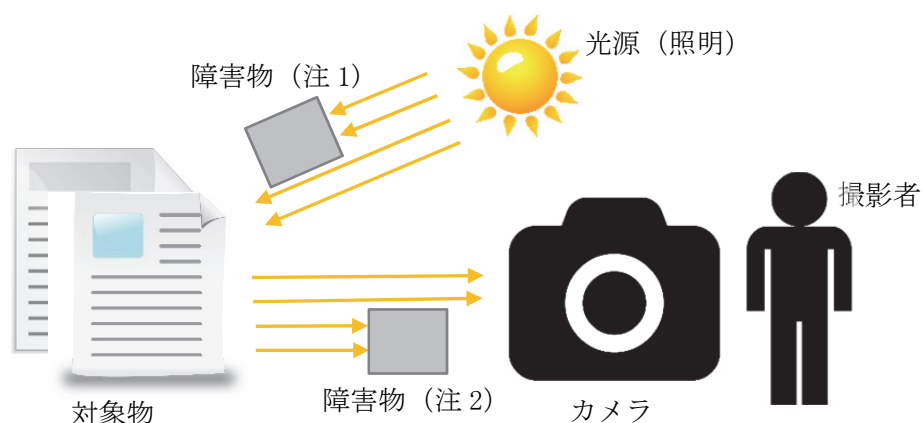
本章では、まず、2.1節にて認識精度に影響を及ぼす外乱要因について現状の整理状況を報告する。2.2節ではOCRやRPAの文脈を中心としながら「デジタル化」というキーワードが指すものについて俯瞰的視点から考察する。2.3節では特に近年注目されているRPAにおいて、UI操作を自動化するためのOCRにおける外乱要因について述べる。2.4節ではRPA向けOCRをはじめとした、実世界環境における認識機器の最新動向と今後の展望について述べる。

2.1 実世界環境における外乱要因

本節では 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述べる。外乱要因の整理はブレインストーミング的に以下の手順で進めている。

- (ステップ 1) 文字認識の専門家視点での外乱要因及び画像劣化現象の網羅的な列挙
- (ステップ 2) それらの見直しと外乱要因の発生源起点でのグループ化
- (ステップ 3) 画像劣化起点でのグループ化
- (ステップ 4) 画像認識の非専門家向けの整理

今回の報告はステップ 2 まで実施した途中結果である。図 2.1-1 に示すように、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は(a)光源、(b)障害物、(c)対象物、(d)カメラ、(e)撮影者が関わり、外乱要因はこれらに起因して発生すると考えられる。そこで、外乱要因をその発生源毎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との関係を整理した。整理した結果を表 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発生する画像劣化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タイプの画像劣化に関係する場合もある。



（注1）影を引き起こす障害物
（注2）遮蔽を引き起こす障害物

図 2.1-1 文書・帳票画像のデジタル化の過程

表 2.1-1 OCR に影響を与える現象及び要因

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 ・ 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-08, L-09, L-10 ・ 低コントラスト L-01, L-06, L-07, L-10 ・ 縞々 L-04, L-10 	<ul style="list-style-type: none"> ・ 状態及び種類 <ul style="list-style-type: none"> - 明るさ (L-01) - 指向性 (L-02) - 点光源/面光源 (L-03) - 周波数 (フリッカー) (L-04) - フラッシュ (L-05) - 光源色/スペクトラム (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 (L-07) - 光源の数 (L-08) - 複数の位置関係 (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)
障害物	<ul style="list-style-type: none"> ・ 像がゆがむ 0-03, 0-04, 0-05, 0-06, 0-08 ・ 影がおきる 0-01 ・ 遮蔽される 0-02, 0-06 ・ ぼける 0-06, 0-05, 0-07 ・ 低コントラスト 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 <ul style="list-style-type: none"> - 空気 (0-03) - 水 (0-04) - ガラス (0-11) ・ 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)

分類	現象	要因
対象物	<ul style="list-style-type: none"> 汚れている T-07 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 ノイズ(画質) ゆがんでいる T-05, T-06, T-21, T-22, T-23 遮蔽される 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 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 文字列を検出しづらい T-28, T-29, T-30, T-32, T-33, T-48 類似文字 T-46 隠れている文字 T-26 	<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 -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 -立体 (T-05) -曲面(皺、波うち、折れ) (T-06) -汚れ・模様 -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"> -色が一様でない (T-12) -印字不良 -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) -濃度が一様でない(T-18) -経年劣化 -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) <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 -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 -直線状にならんでいない文字 -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)

分類	現象	要因
		<ul style="list-style-type: none"> -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）
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-13, C-17, C-18, C-19, C-30 ・白飛び C-03, C-06, C-12, C-13, C-15, C-17, C-18, C-19, C-30 ・ピンぼけ C-03, C-08, C-09, C-29 ・二重（多重）にみえる C-07, C-09, C-23 ・ゆがむ C-02, C-10, C-16, C-27 ・低コントラスト C-04, C-06, C-12, C-13, C-17, C-18 ・低解像度 C-01, C-02, C-11, C-26, C-31 ・ノイズ（ホワイトノイズ、縞々ノイズ、ブロックノイズ） C-14, C-15, C-20, C-21, C-24, C-25 	<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"> -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 -減光（C-04） -解像度低下（C-23） -F値（C-06） -内面反射（C-07） -被写界深度（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） -感度（ISO）（C-13） -暗電流ノイズ（C-14） -スミア（C-15） -ローリングシャッター（C-16）

分類	現象	要因
		<ul style="list-style-type: none"> ・機構系 <ul style="list-style-type: none"> -露出 (C-17) -EV 値 (C-18) -絞り (C-19) -シャッター速度 (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) -ベイヤール配列 (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 -ピント (C-29) -露出 (C-30) -ズーム (C-31)

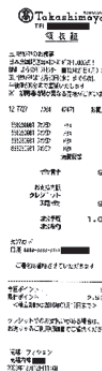
(注) 青字は従来型スキャナーを入力装置とする OCR にも共通する外乱要因を示す

なお、従来型のスキャナーを入力装置とする OCR における外乱要因はこの表に網羅される。具体的には、(c)に起因する青字で示した外乱要因が従来型のスキャナーを入力装置とする OCR にも共通する外乱要因となる（一部例外あり）。

また、本年度の委員会では、主要な外乱要因についてその具体例と上記整理への網羅性を再確認するために、対象を文書・帳票以外の文字情報にも広げて、OCR 読み取り困難だった事例の収集を行った。挙げられた読み取り困難事例の中からいくつかの例を図 2.1-2 に示す。この作業を通じて、各事例に対して OCR 読み取りを困難にする要因の大部分は、もともとの外乱要因リストに網羅されていることを確認した。また、外乱要因リストに網羅されていなかった一部の要因については表 2.1-1 に反映済みである。



(a) 書籍や雑誌の表紙画像



(b) レシート



(c) ゼッケン



(d) ナンバープレート

(出典) (a) 電子情報通信学会 情報・システムソサエティ誌
 (c) <https://www.sigfoss.com/casestudy/number-recognition>
 (d) <https://www.parking-s.co.jp/buiness/numberplate.html>

図 2.1-2 OCR 読み取り困難事例の一例

今後は、外乱要因によって生じる現象中心での整理、及び画像認識の非専門家向けに主要な外乱要因の選定及び外乱要因の詳細化を施した整理を進めていく予定である。

2.2 デジタル化に関する俯瞰的考察

社会のデジタル化は急速に進んでいるところであり、政府も 2021 年 9 月にデジタル庁を設置するなどして推進体制を積極的に強化している。また、2020 年 4 月に出された新型コロナウイルス感染症による緊急事態宣言から続く、一連の感染防止対策によって会議やイベントのオンライン化が急務となったほか、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た。

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ると捉えることができ、そのような意味ではデジタル化の先駆けと考えることもできる。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。)

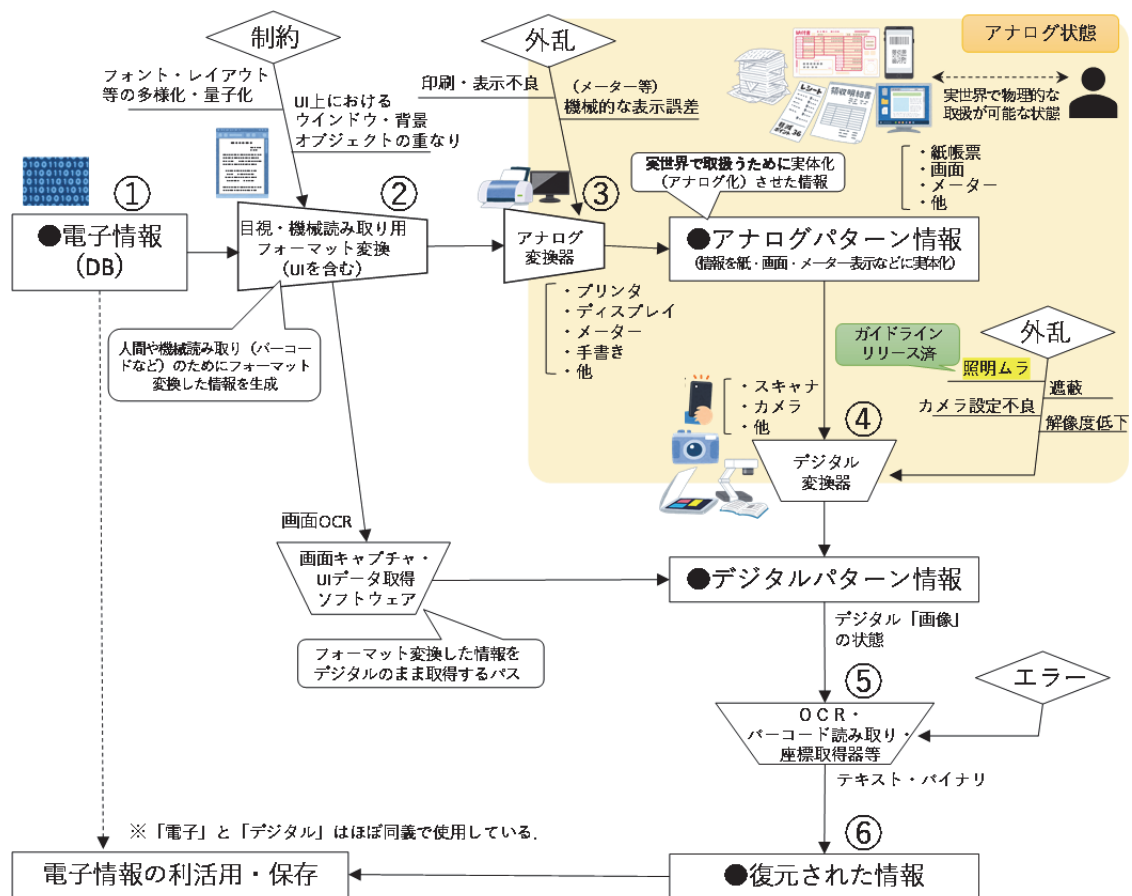


図 2.2-1 デジタル化に関する俯瞰的な考察

図 2.2-1 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、通常の人間用の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。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。

2.3 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 2.3-1）。つまり完全な自動処理ではなく、認識精度は人が責任を持つ必要がある。それでもキーボードでデータを手入力するよりは作業工数が大幅に削減できる。若干の誤りであれ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 2.3-2）。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 2.3-1 RPA 向けデータ入力自動化のための OCR 利用



図 2.3-2 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。

2.3.1 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レコーディング（操作記録）機能を提供している。（図 2.3-3）

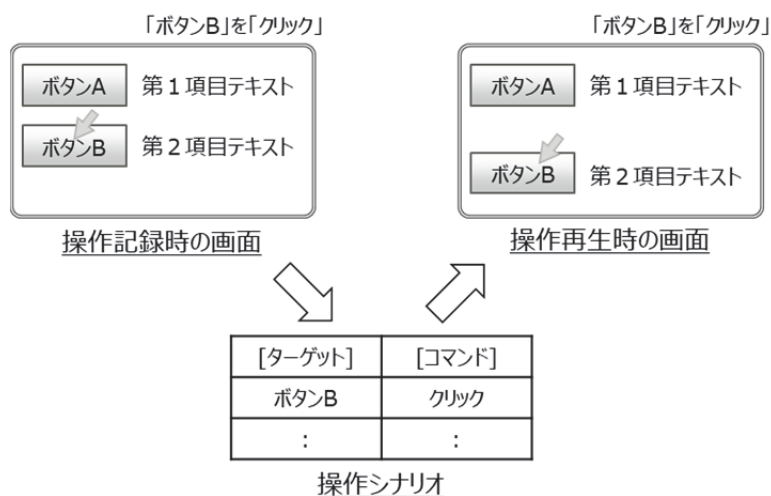


図 2.3-3 RPA における UI 操作の記録と再生

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 2.3-4 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 2.3-5 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。

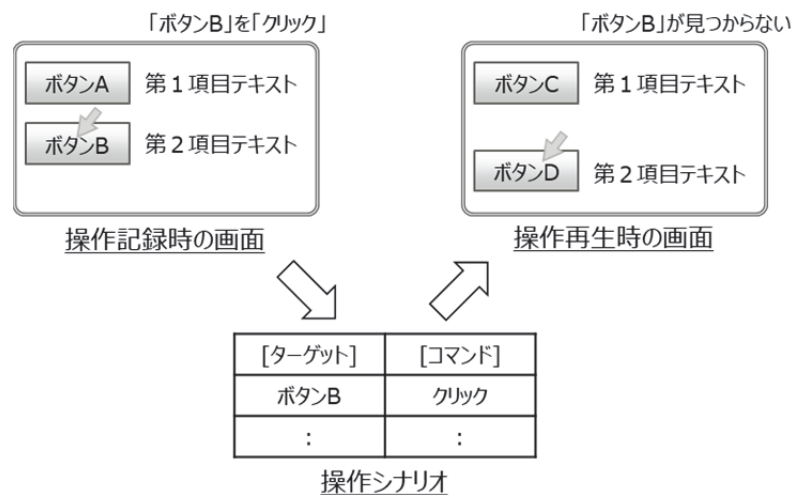


図 2.3-4 RPA における UI 操作の記録と再生（失敗例）

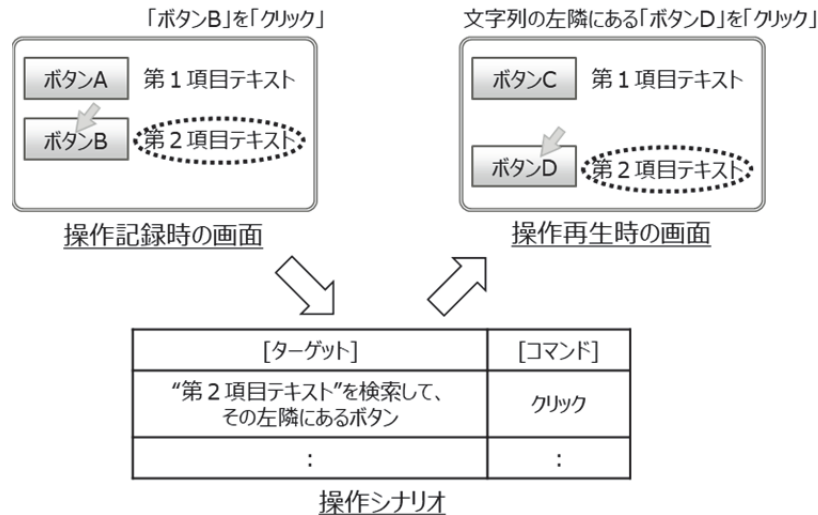


図 2.3-5 RPA における UI 操作の記録と再生（OCR の利用）

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)

2.3.2 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 2.3-6 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、OCR の精度に影響が生ずる。画面 OCR の実行条件を設定する際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 2.3-6 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変化も生じ得る。画面 OCR の精度を維持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 2.3-1）。

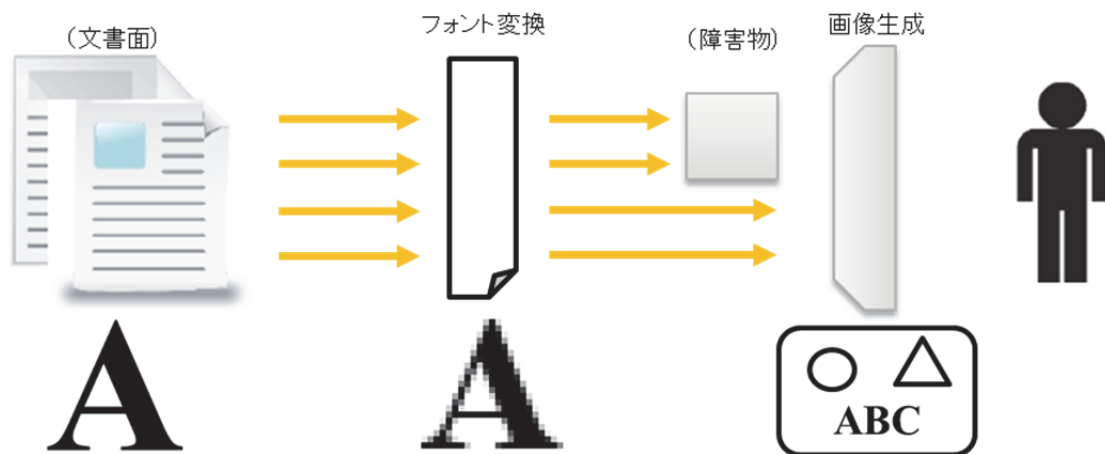


図 2.3-6 画面キャプチャー画像の撮像過程に影響を与える要因

表 2.3-1 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）、画面解像度

2.3.3 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 2.3-1）が、RPA 実行時に OCR を行う UI 操作自動化においては（図 2.3-2）、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る。しかし、現実には適切な品質保証戦略は確立しておらず、しばしばシステム導入の障害となっている。

パターン認識の品質問題は、近年「AI 品質」の問題として議論が活発化している。この問題はソフトウェアエンジニアリングの分野で議論されることが多く、文書認識の技術分野ではあまり話題に上っていない。しかし、誤動作による損失が大きな用途に OCR を用いようとするとき避けられない問題であり、文書認識でも品質保証問題の議論が活発化することが期待される。そこ

で、OCR を RPA（自動実行）で活用する際に必要な情報として、AI 品質に関する国内の主要なガイドラインを下記に記す。

■ 機械学習品質マネジメントガイドライン 第2版（通称：産総研 AI 品質ガイドライン）

<https://www.digiarc.aist.go.jp/publication/aiqm/guideline-rev2.html>

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。企業・大学などの外部有識者とともにまとめたもの。

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ガイドライン。

■ 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/>

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 2021 年 9 月版が最新である。

2.4 実世界環境における認識機器の現状と今後の展望

2.1 節に説明したように、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続けているが、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められてきた。

本章では、2.4.1 で、OCR と技術的に関連している音声認識や自然言語処理技術の最近のトピックスを紹介し、2.4.2 では「OCR 技術の進化と、OCR における実世界環境対応について」、2.4.3 では「DL で変化した文字・文書認識技術」、2.4.4 で「今後の展望」について記述した。

特に、2.4.2 では OCR 商品の歴史（概要版）や、実世界環境対応の推移（概要版）について、簡単な年表を作成した。また、AI-OCR に関しては、2.4.3 で「AI-OCR の特長」、「AI-OCR で認識精度が向上した文字認識技術について」、「色々なところで利用され始めている AI-OCR」について最近の動向を記述した。

2.4.1 実世界環境における認識機器（技術）のトピックス

最近の認識機器（技術）は、DL などの AI 的な手法を積極的に導入することで、実世界環境に

おける認識精度が向上し、色々な分野で利用されるようになってきた。トピックスとして、特に進化が著しい音声認識技術と自然言語処理技術について紹介する。この2つの技術は、1960年代から研究が盛んに行われてきたが、2.4.2以降で紹介する文字・文書認識技術とも密接な関係があり、互いの研究成果を利用している場合も多い。

1) 音声認識技術

DLを利用して大量の学習データを End-to-End (E2E) で学習することにより、実世界環境における認識精度が急速に向上しており、スマートスピーカー (AI スピーカー) という音声認識機器 (技術) が色々な場面で利用され始めている[1]。スマートスピーカーは、長年、制限環境下で利用されてきた音声認識・対話技術を実世界に近い環境で利用できるようにした技術で、音声検索や質問・応答、家電機器の音声操作などに利用され始めている。

コンピューターが音声を介して人と対話しながら Web 上の知識にアクセスし、情報の検索を支援するソフトウェアエージェントの考え方は、ジョン・スカリー (Apple 社の元 CEO) によって 1987 年に提案された (Knowledge Navigator) が、このデモ動画[2]を見て、当時の国内における音声認識研究者は Apple 社の目標の高さに刺激を受けた。このデモ動画の技術的なポイントは、①タブレット端末、②音声認識技術と音声による対話、③情報検索、④テレビ電話、⑤スケジューリング管理、などであるが、これらの機能は、現在のスマートデバイス (スマートフォンやタブレット端末) で、ほぼ実現可能になっている。

実世界環境における音声認識精度の向上は現在も継続されており、代表的なヒューマンインタフェースの1つとして、今後も発展していくと予想される。

2) 自然言語処理技術

自然言語処理技術はここ2~3年の間に激変し、対話エンジンの基盤技術が、「Transformer」モデルなどにより「飛躍的」に向上した。そして、対話エンジンの基となる DL の「言語モデル」では、モデルの規模が大きければ大きいほど質も同時に高まる、「スケーリング則」が成り立つことが経験的に知られている[3]。2020年5月に米 OpenAI が発表した「GPT-3」から、世界中で言語モデルの大規模化競争が始まり、GPT-3 のモデル規模は 1750 億パラメータだった。また、2020年6月には米グーグルが 6000 億パラメータの「GShard」を発表し、2021年1月には 1 兆 5710 億パラメータの「Switch Transformer」を発表した。さらに、2021年6月には、中国の北京智源人工智能研究院が 1 兆 7500 億パラメータの「Wu Dao」を発表した。

わずか1年ほどの間に、言語モデルの規模は 10 倍になり、モデルの大規模化も進んでいる。大規模な言語モデルを生成するには1回の学習だけで数億円以上の投資が必要といわれており、自然言語処理向けの AI 技術は、世界の巨大企業の投資競争になりつつある。

2021年11月、日本語においても LINE 社が、元・親会社の韓国 NAVER 社と共同で 390

億パラメータの巨大な言語モデル「Hyper CLOVA」を日本語ベースで開発している。

これらの成果は、文章・画像生成や翻訳、文章からプログラミング言語「Power Fx」への自動変換[4]、などにも利用され、人間に迫る成果が出始めている。巨大な言語モデルは、文字・文書認識分野においても、誤読文字の確認（発見）・修正、認識精度の向上、などに利用できると考えられ、今後に期待したい。

上記1）、2）に示したように、画像認識を得意とする DL 技術の出現は、これまで難しかった認識機器（技術）の実世界環境での使用を可能にしつつあり、頑健性や公平性、透明性などでの問題が指摘されてはいるものの、色々な分野で利用され、成果を出し続けている。

2.4.2 OCR 技術の進化と、OCR における実世界環境対応について

2.4.1 で実世界対応が進んでいる認識技術を2つ紹介したが、本委員会で長年、調査検討を進めてきた OCR 関連技術（文字・文書認識技術）においても、「認識対象となる文書や帳票」及び「入力画像」において実世界環境対応が進んでいる。本節では、OCR の歴史を簡単に振り返りながら、その時々に関与しようとした課題、及び、その解決手法などについて紹介する。

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、色々な分野で利用されている[5]。「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）して、誤読を最小限に抑える工夫をしながら、データの入力ツールとして長年利用されてきた。一方、「文書 OCR」は、新聞、雑誌、教科書、名刺、などを読み取ることができる OCR で、読み取り文書のレイアウト解析を行い、テキスト領域における色々な書体の活字文字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」ともに、入力画像や読み取り対象文書・帳票に対する実世界環境対応への要望は多く、研究開発は継続的に進められてきた。しかし、デジカメで撮影した画像の読取りや、非定型帳票の解析・理解、フリーピッチで書かれた手書き文字列の高精度認識、一般文書のレイアウト解析・理解などは難しいテーマであり、整備（制限）された環境内で使用されてきた。

当初、整備環境下で利用されていた OCR であるが、計算機環境（CPU、メモリ、開発環境、GUI、など）の進化に伴い、徐々に実世界環境対応も進んでいった。図 2.4-1 は、これまでにリリースされた色々な OCR 商品を時系列に表示したものである。

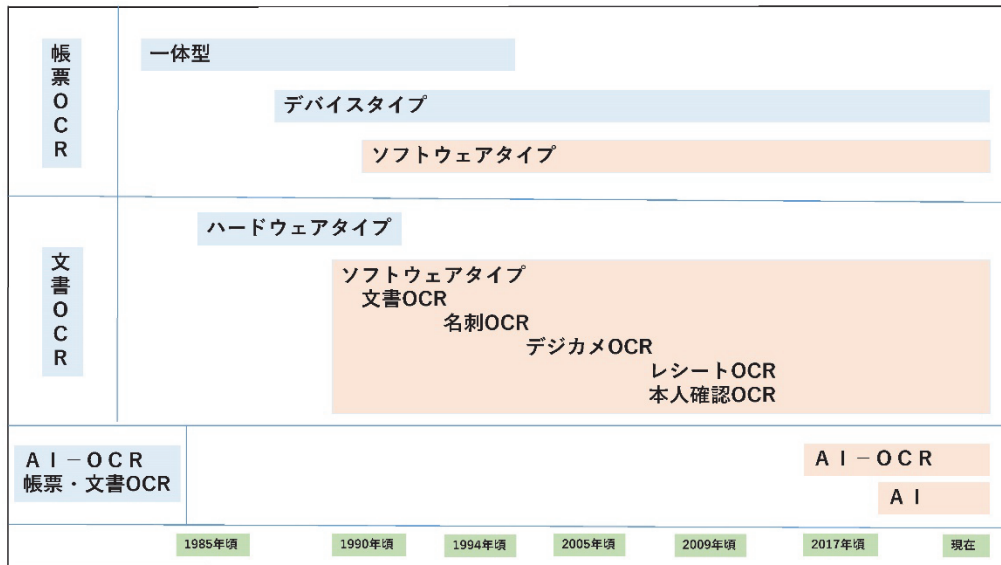


図 2.4-1 OCR 商品の進化（概要）

「帳票 OCR」は、1980 年代は、ハードウェア中心の一体型であったが、徐々にスキャナ一部分が分離したデバイスタイプになった。1971 年に Intel 社が 4 ビット CPU をリリースし、その後、Intel 社と Motorola 社を中心に CPU が急速な進化を続け、PC の OS も Window 環境 (Mac、Windows、Unix、他) での操作が主流になっていった。

OCR は当時、他の制御ソフトと比較すると、動作させるために多くのメモリ空間が必要なことや認識結果の確認・修正を行う GUI として Window 環境が適していたことなどから、1991 年に、PC の Window 環境で動作する日本語文書 OCR ソフトが国内で初めてリリースされ、その後、OCR のソフトウェア化が加速していった。CPU の高速化やメモリの大容量化、プログラム環境なども継続して進化していき、認識部分の研究開発が PC 上で可能となり、活字文書 OCR、名刺 OCR、デジカメ OCR、レシート OCR、本人確認 OCR、などが、PC のソフトウェアとして、次々とリリースされた。

また、デジカメ OCR が商品化された 2005 年頃から、スキャナー入力が主体であった OCR は、デジカメ（デジタルカメラ、スマートデバイスのカメラ、など）入力への対応が進み、入力画像に対して実世界環境（デジカメ画像）に対応した色々な OCR（名刺、免許証、保険証、マイナンバーカード、パスポート、など）が商品化された。この背景には、スマートデバイスの急速な進化と普及があった。

また、2017 年ころから、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普及により、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、という認識対象に対しても、実世界環境対応への要望が多くなり、これらに対応できる高精度な OCR の開発が課題になった。

一方、2012 年に ILSVRC (Imagenet Large Scale Visual Recognition Challenge) において、

DL を利用して高い認識精度を達成したという発表があり、その後、DL は色々な分野で高い認識精度を達成してきた。2016 年頃に、文字・文書認識の分野においても、DL を利用した高精度な手書き数字認識の発表（九州大）などがあり、これまで難しいと考えられてきた課題が DL で解決できる可能性が見え出した。特に、3 章で認識技術の動向として報告してきたように、「非定型帳票」の解析や、「フリーピッチ手書き文字列」の認識などでは、DL を利用した方式が従来方式を上回る結果が報告され、その成果は 2017 年ころに「AI-OCR（AI という場合もある）」という名称で商品化され始めた。最近では、従来 OCR に DL 技術を搭載して精度向上した AI-OCR や、ベンチャー企業の開発した AI-OCR が多数リリースされ、OCR 市場が活性化している。AI-OCR が出現した背景は、ユーザー要望に対して、DL 技術を利用した認識技術の進化と、計算機環境（CPU/GPU の高速化など）の進化、などが併行して進んだ結果とも考えられる。

図 2.4-2 は帳票 OCR、文書 OCR、AI-OCR における実世界環境への対応状況（概要）を示したものである。OCR の実世界環境対応を、①「非定型帳票」の認識、②「フリーエリアの手書き文字列」の高精度認識、③文書・帳票のレイアウト解析及び認識、④「デジカメ画像」の認識、という課題ととらえた場合、これらの課題は、帳票 OCR、文書 OCR、AI-OCR のリリースによって徐々に進化してきていることがわかる。

DL は、2.4.1 で説明したように、色々な分野で成果を出しているが、文字・文書認識の分野においても、特に実世界環境対応を進めていくための強力な手法の 1 つとして期待されている。

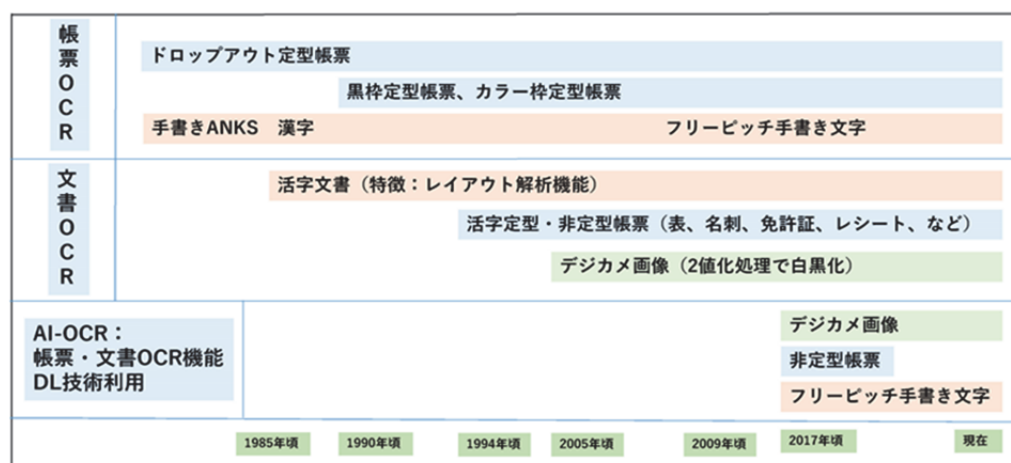


図 2.4-2 OCR 商品の実世界環境への対応状況（概要）

2.4.3 DL で変化した文字・文書認識技術

1) AI-OCR の特長

「働き方改革」を進める中で、PC 作業の効率化を目的に RPA 市場が拡大している。RPA を導入する場合に、OCR を利用して読みたい文書や帳票は、経理、医療/介護福祉、学校関

係、交通チケット、カード、などの「非定型帳票」や、「フリーエリアに書かれた手書き文字列」などが対象になる場合が増えている。

帳票 OCR では、帳票矩形の位置情報や、その属性を帳票定義として用意し、帳票定義の情報に従って帳票内のデータを属性付きで読み取る方法がとられてきた。当初は、帳票自体も、文字抽出を簡素化するためにドロップアウトカラーで印刷していたが、その後、黒枠・カラー帳票を読める機能が搭載され、普通に印刷された帳票も読めるようになった。しかし、使用される帳票はタスクごとに決まったフォーマットになっている定型帳票が主流であった。RPA で要望される非定型帳票を読む OCR では、色々な会社の伝票などを、帳票枠を検出して、帳票の構造を理解し、読んだデータを帳票属性どおりに出力する必要がある、この仕組みを従来技術で作るのは難しかった。

2017 年頃、従来 OCR の課題であった、①「非定型帳票」の認識、②「フリーエリアの手書き文字列」の高精度認識、③文書・帳票のレイアウトの解析、④「デジカメ画像」の認識、などを、DL を利用して高精度化した OCR が商品化されるようになり、「AI-OCR」と呼ばれるようになった。

AI-OCR は、DL を使用して精度向上していることから、従来 OCR と比較して計算量が多くなる傾向があり、GPU を搭載した高速 PC での動作の推奨や、クラウドサービスで提供される場合も多い。また、AI-OCR では、従来の「帳票を読む機能」と「文書を読む機能」の両方を備えている場合が多く、「帳票 OCR」と「文書 OCR」の区別は無くなってきている。

2) AI-OCR で認識精度が向上した文字認識技術について

・非定型帳票の認識

昨年度の報告書に DL を利用した領収書認識の一例を示したが、この例では認識領域を検出し、検出した領域を文字認識して結果を出力するという方法がとられている。従来は、まず、2 値化してから文字領域を抽出・認識していたが、DL を利用した場合、カラー画像から直接文字列領域を抽出することも可能になりつつある。詳細については、2020 年度の当委員会報告書[6]を参照願いたい。

・フリーピッチ手書き文字列の認識手法例

フリーピッチの手書き文字列の認識は、従来手法においても、手書き数字に関しては人間に近い認識精度になりつつあったが、漢字を含めた手書き文字列に関しては、住所・氏名などの限定した範囲での高精度化が進められてきた。従来 OCR における、漢字を含む手書き文字列の認識は、一般に文字画像片を矩形として抽出し、行や文章単位で予測した文字幅を参考にして、文字矩形画像を認識しながら、文字矩形画像片どうしを、知識処理を併用しながら、結合と分離を繰り返して認識していく、という手法がとられている場合が多い。

DL を利用する場合、昨年度の当委員会報告書[6]に一例を示したが、その方式とは別に、

図 2.4-3 に示したように、文字列画像を一定間隔で切り出し、切り出した文字片画像を CNN (Convolutional Neural Network) で認識し、その結果を RNN (Recurrent Neural Network) や LSTM (Long Short-Term Memory) に入力し、文字片画像単位に出力された認識候補文字列を、CTC (Connectionist Temporal Classification) decoder を利用して、最終的な文字列に変換する、という E2E な方式が提案されている[7]。

図 2.4-3 の方式は、連続音声認識で使われている手法と考え方は同じであり、DL を利用して、フリーピッチの手書き文字列も、これまで以上の精度で認識することが可能になりつつある。この手法は、通信学会で行われた「くずし字」認識コンテストにおいても高い評価を得ていた。

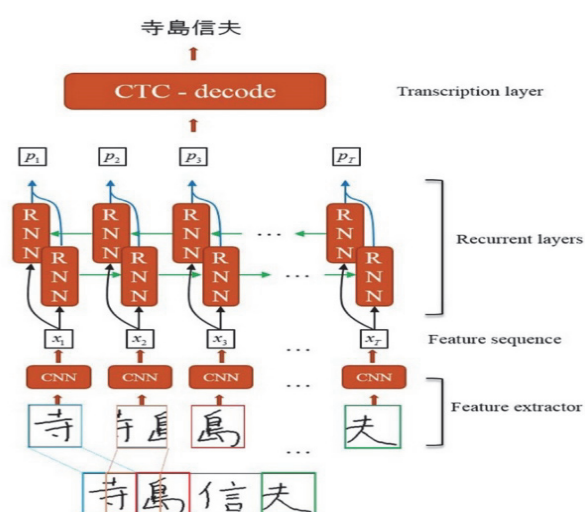


図 2.4-3 フリーピッチの手書き文字認識方式例[7]

・デジカメ画像の認識（情景画像中の文字認識）

ビデオカメラを含むカメラ画像を認識する研究開発は、1990 年代後半から盛んになり、スマートデバイスの普及とともに、色々な商品がリリースされた。当初は、入力画像を 2 値化してから従来 OCR に入力し、認識する方式が主流であった。そこで当時は、色々な実世界環境で撮影されたカメラ画像を確実に 2 値化することが研究課題の 1 つになっていた。

また、カメラ画像は撮影の仕方によって、2.1 節で議論している、ボケや変形、背景雑音などが加わるので、ボケや変形した画像の補正なども研究された。しかし、これらの課題を従来技術ですべて解決することは難しく、使用方法を制限する場合が多かった。

図 2.4-4 は、2001 年頃に実用化を考えていた「ハンドマウスの応用：ユニバーサルリモコンと実環境中の文字情報取得」という論文の中で紹介されている「実世界環境からの文字情報取得」を行う例である。画像中に存在する文字列を自動で検出するのは、背景模様などから難しい場合が多く、文字の存在するエリア付近を指で示し（論文ではハンドマウスと呼ん

でいる）、その付近を中心に文字列を探して認識するという方式をとっていた[7]。



図 2.4-4 実世界環境からの文字情報を取得する例[7]

その後、情景画像中の文字認識においても、DL を利用して、文字列の検出や文字認識などを行う手法の研究が進んでおり、ゼッケン番号の認識や、缶に印刷されている文字列の認識、標識の認識[9]、などの研究成果が発表されている。ゼッケン番号の認識は、各地で行われているマラソン大会などで、自分が走っている写真をゼッケンナン番号で探して購入できるサービスなどで実際に利用されている[10]。

3) 色々なところで利用され始めている AI-OCR

- ・ RPA における文字画像データの読取り

RPA において、画像（帳票画像や PC の画面画像など）を、OCR や AI-OCR を利用してテキスト化して入力する方法は徐々に普及しており、データ入力の効率化やシステムの自動化に貢献している。

OCR の RPA での利用に関しては、2.3 節や、2019 年度の本委員会報告書[11]で紹介しているので参照いただきたい。

- ・ eKYC（オンライン本人確認）での利用

従来の対面／郵送での本人確認は KYC（Know Your Customer）と言われている。これに対して、オンライン上で本人確認を行う技術は、オンライン上で行うという意味で e（electronic）という文字を追加して、eKYC（electronic Know Your Customer）と呼ばれている。

eKYC では、免許証や保険証、マイナンバーカード、パスポートなどをスマートデバイスのカメラで撮影して転送されてきた画像を認識することが多い。カメラで撮影した画像を読み取る場合、2.1 節で検討している「実世界環境における外乱要因」が課題になるが、DL を利用して撮影環境の影響を少なくした AI-OCR などを使用し、運用制限やセキュリティ面での対応を行い、eKYC での利用が進もうとしている。

2.4.4 今後の展望

2.1 節の文字認識における「実世界環境における外乱要因」に示したように、現状では色々な環

境の中で文字を正確に認識するためには課題も多い。

AI-OCR では、これまで実現するのが難しかった機能や認識精度に関する課題を、DL を利用して少しずつ解決しつつある。DL を利用するためには多くの学習データが必要になる場合が多く、システムを運用しながら学習データを収集する方法や、データ拡張方式を利用する方法が提案されている。また、DL を利用した精度向上では、大量の学習データを巨大なモデルで継続的に学習することが必要との指摘もあり、精度向上は計算機環境に依存（計算機環境を持つものが有利）するとも言われ始めている。

OCR における「実世界環境における外乱要因」に関連する課題は、DL や次の新技術を利用して、計算機環境の進化とともに解決していくと予想されるが、機能・性能の向上に伴い、OCR の新しい市場がさらに拓けていくと考えられる。入力画像や認識対象の実世界環境対応に向けて、AI ベンチャーや従来からの OCR メーカーが DL を利用した AI-OCR の機能・精度向上を進めており、ユーザーも拡大しているので、今後の展開に期待したい。

当委員会では、認識機器の実世界環境対応に向けて、認識精度に影響を及ぼす「実世界環境における外乱要因」に関する調査・検討を継続している。この調査は今後、実世界環境に対応した OCR を研究開発していく過程で、①使用する場合のガイドラインの作成、②実世界対応するための学習データの効率的な収集、及び作成、③成果物の評価、などに利用できると考えている。

今後も、実世界環境対応に向けた認識機器の研究開発が促進していくために、課題や関連技術の調査・検討等を継続し、報告していきたい。

【参考文献（URL は 2022 年 2 月 24 日時点）】

- [1] スマートスピーカー（AI スピーカー）比較と選び方のポイント
<https://kakaku.com/pc/article/smartspeaker/>
- [2] Apple 社のデモ動画（Apple Knowledge Navigator Video (1987)）
<https://www.youtube.com/watch?v=yc8omdv-tBU>
- [3] OpenAI が発見した Scaling Law の秘密
https://deeplearning.hatenablog.com/entry/scaling_law
- [4] 文章をプログラミング言語に変換、Microsoft がローコード開発ツール「Power Apps」に「GPT-3」活用 <https://www.itmedia.co.jp/news/articles/2105/26/news087.html>
- [5] 入力装置に関する調査報告書 JEITA（情産－21－情端－3）2021 年 7 月
- [6] 認識形入力方式に関する調査研究報告書 JEITA（情産－21－技標－1）2021 年 3 月
- [7] Nam -Tuan Ly、他「Deep Convolutional Recurrent Network for Segmentation-Free Offline Handwritten Japanese Text Recognition」ICDAR 2017
- [8] 蔵田、他「ユニバーサルリモコンと実環境中の文字情報取得」INTERACTION 2002
- [9] Fangge Chen、他「Text Detection in Traffic Informatory Signs Using Synthetic Data」ICDAR 2017
- [10] （お知らせ）マチエミとドコモ、画像認識エンジンを活用した新しい写真販売サービスを
開発 https://www.nttdocomo.co.jp/info/news_release/2020/06/01_01.html
- [11] 認識形入力方式に関する調査研究報告書 JEITA（情産－20－技標－1）2020 年 3 月

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、また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、スキャナー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[1]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、十分な学習データを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。

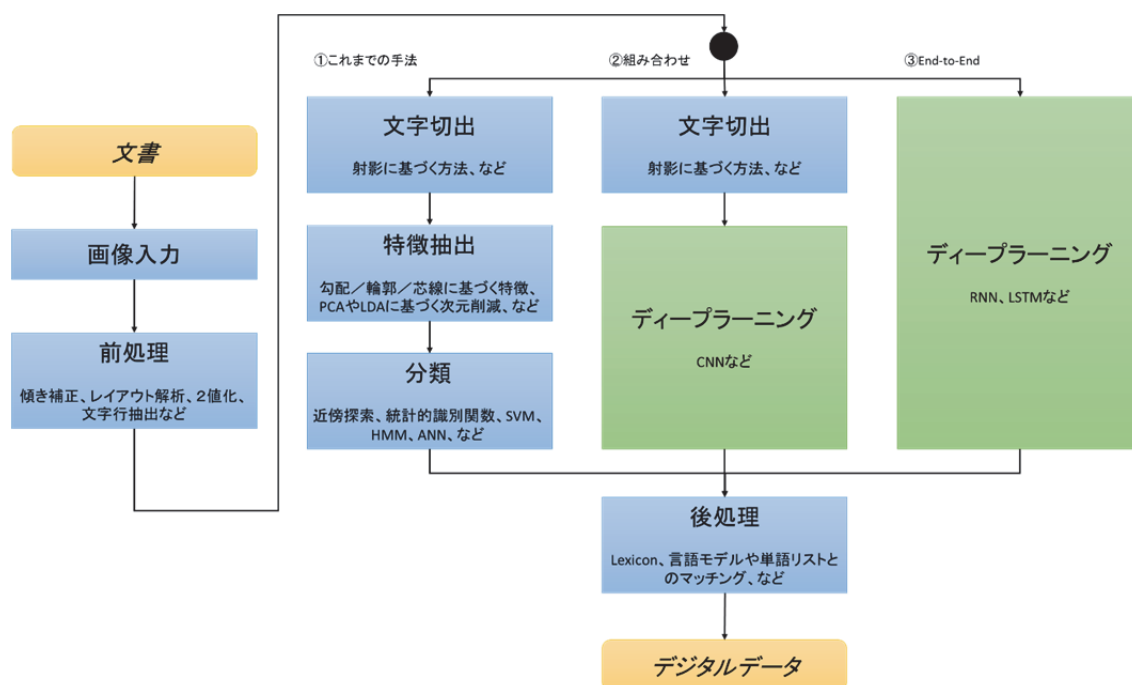


図 3.1-1 文字認識技術の概略

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[2]や Cogent Labs の Tegaki[3]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。

しかし、新たな技術の導入により、これまでの OCR にはなかった新たな課題も生じている。代表的な例として人間の直感に反する誤認識、特に重複読み・読み飛ばしの発生が挙げられる。例えば、「信」という文字の一部（にんべん）を 2 回認識して「イ信」、あるいは読み飛ばして「言」が認識結果となるような事例である。こういった誤認識は画像データを集積し、それらを新たな学習データとして再学習を行うことで軽減可能である。しかし、人間が書く文字は極めて多彩なのですべてを収集することは難しく、学習データの強化だけで完全に防ぐことまでは期待できない。今後はこういった人間にとって違和感のある誤認識を削減する技術開発が期待される。また、多くの利用者に広げていくためには、Explainable AI (XAI) [5]を取り込むなどして認識結果の説明性向上を図り、利用者にとって理解、納得のしやすい文字認識へと発展していくことが重要と考えられる。

また、サービスの普及に伴って品質保証の議論も活発化している。既に QA4AI[6]に AI-OCR についての記述が加えられており、今後はこういった観点での説明も利用者にとって重要になってくるものと思われる。

全体的には AI-OCR の認知・理解が堅調に進んでおり、今後ますます普及していくことが期待される。

3.1.3 RPA での OCR 利用

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ぼ同一の帳票、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。このような OCR の代表的な例として Infordio の smartOCR[6]がある。

こういったサービスでは、項目に付与された題字や項目自体の書式、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 OCR（いわゆる定型帳票認識）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。このような OCR もまた、AI-OCR と呼ばれ、これまでは請求書などの比較的単純で多くの枚数が扱われる帳票が対象となっていたが、技術開発が進んだことで対象となる帳票の種類が拡大しており、ガソリンスタンドのレシート[9]、処方箋[11]といった帳票を扱うものが出てきている。

こういったサービスでは、読取結果に対するユーザーの訂正などに基づき、運用中に項目の位置、項目同士の位置関係を学習したり、あるいはユーザーがキーワード登録などのカスタマイズを行ったりすることで精度向上を図れる場合も多い。このような利用するほどユーザーに合わせて「賢く」なっていく学習機能は AI-OCR にとって重要な機能になっている。このような観点での新たな取り組みとして、アライズイノベーション社が、同社の AIRead にオンプレミス環境での学習機能を搭載したことを発表している[13]。クラウド環境では学習機能は広く利用されているが、OCR で読み取る帳票は個人情報など機微な情報を含むものも多く、学習のためにデータをインターネット上にアップロードすることは多くの場合望ましくない。データ数が相対的に不足するという課題はあるが、今後はオンプレミス環境での学習機能も進展していくものと思われる。

RPA 用途において特に注意が必要となるのは、OCR の認識率が 100%になることはなく誤認識は必ず発生するという点である。既存の OCR システムでも必ず誤認識は発生していたが、人間による訂正確認フェーズを業務フローに組み込むことでこの課題を補ってきた。RPA 連携では認識結果を確認・訂正することなくそのまま次の処理で利用する場合があります、この点が大きな差異である。このため、誤認識の発生を前提とし、それでも問題が発生しないようなシステム全体の設計・運用を行うことが求められる。

3.1.4 AI-OCR 活用の拡がり

2020 年から続く新型コロナウイルス感染症（COVID-19）の問題により、生活様式の大きな転換が進んだ。

会社業務においてはテレワークが積極的に推進されている。2020 年に比べるとテレワークの比率が低くなってきてはいるものの、コロナ以前との比較ではその比率は大幅に増加しており、今後も無視できない規模で継続されることが予想される。

テレワークの継続・拡大のための課題としては「社内の打合せや意思決定の仕方の改善」、「書類のやりとりを電子化、ペーパーレス化」等が引き続き挙げられている。企業の業務では他社やユーザーとのやり取りが多く発生し、これらのやり取りを完全に電子システム上で行うことは難しいため、紙、ないし紙をスキャンした（電子化されたテキストを含まない）データが送られてくるケースが依然として多いことや、紙を用いた業務になじんでいる人にとって電子システムへ

の移行は難しい場合があることなどから現状では紙を用いた業務が多く残っており、また、これらがすぐになくなることも考え難い。これらの業務で用いられる「紙」に書かれている情報を電子化してしまえば、これらの業務もオンラインで実施可能であり、これに AI-OCR を用いるニーズがある。また、受け取った電子データに電子化されたテキストが埋め込まれている場合であっても、単純にテキストが埋め込まれているだけで構造化されておらず、そのままでは業務に有効活用できない場合も多い。このような場合に AI-OCR の項目抽出技術を用いて埋め込まれたテキストデータを構造化する需要が存在する。今後、このように AI-OCR を構成する技術の一部を活用する応用もより注目を集めると思われる。

AI-OCR の活用事例に対するニーズが高まっていることから情報共有が進んでおり、例えば、アイスマイリーが業界別に AI-OCR ソリューションの活用事例をまとめた「AI-OCR 業界別事例集」を公開している[15]。また、兼松エレクトロニクスの提供している定型業務改善ソリューション RWR (Routine Work Reformer) [17]のように、AI-OCR 自体をセールスポイントとするのではなく、より上位の視点で効果を訴求し、必要な機能の一部として AI-OCR を含む事例も今後増えてくるものと思われる。

また、OCR 技術が進化したことで新たな活用事例も増えてきている。帳票以外での OCR 技術活用としては Digital Transformation (DX) の一環としてメータ表示などを認識する例が代表的だが、それ以外の活用事例としては、例えば、TDAI Lab がテレビ朝日と共同で、スポーツの生放送にてランキング内の選手名 CG を「英語表記」から「日本語表記」へ書き換える実験を行い、実用性を示したとしている[18]例がある。テロップ認識は長年研究されてきた応用事例であるが、文字の背景となる映像、用いられるフォントなどが多岐にわたり、かつ複雑なものが多いため、正確かつ高速な処理は難しいとされていた。今回の事例ではリアルタイムで翻訳まで含めて処理できたとしており、今後の活用が期待される。

【参考文献】

- [1] Gated Recurrent Convolution Neural Network for OCR J. Wang, NIPS2017.
- [2] DX Suite デジタル変革を加速する <https://inside.ai/dx-suite/>
- [3] Tegaki | 手書き文字をデータ化する AI OCR
- [4] <https://www.tegaki.ai/>
- [5] Explainable Artificial Intelligence (XAI)
<https://www.darpa.mil/program/explainable-artificial-intelligence>
- [6] AI プロダクト品質保証コンソーシアム
- [7] <http://www.qa4ai.jp/>

- [8] スマート OCR | AI-OCR で非定型帳票（請求書・発注書など）をデータ化
<https://www.smartocr.jp/>
- [9] ガソリンスタンドのレシートの文字を読み取り、油種・給油量・給油金額を解析する OCR
ライブラリ提供開始
- [10] <https://www.isp21.co.jp/news/3581/>
- [11] 東大発クラウド処方箋読取サービス「mediLab AI」が AI 学習モデルを一新し、読み取り
精度を改善
- [12] <https://prtimes.jp/main/html/rd/p/0000000005.000066671.html>
- [13] アライズイノベーション、AI OCR「AIRead」の新バージョン 3.2 をリリース
- [14] <https://airead.ai/news/20211214/>
- [15] AI ポータルメディア「AIsmiley」、Web 雑誌「AI-OCR 業界別事例集」を無料配布！
- [16] <https://prtimes.jp/main/html/rd/p/0000000061.000053344.html>
- [17] 定型業務改善ソリューション(RWR)
https://www.kel.co.jp/service/industries/routine_work_reformerrwr_-_rwr_fax.html
- [18] スポーツ生放送中の AI を用いた選手テロップの日本語化
<https://prtimes.jp/main/html/rd/p/0000000011.000050703.html>

(URL 確認 2022 年 2 月 22 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 2021 年 1 月から 2021 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 5 件あった[1][2][6][7][23]。

発表[1]では、CNN を用いた手書き楽譜画像の音楽記号の分類において、小規模のデータセットの転移学習で高い精度が得られることを確認した。約 1.5 万件のデータセットを用いた 7 層 CNN の正解率が 93.1%に対し、データセットの内約 200 件のみを用いて事前学習済ネットワーク Inception-V3 の転移学習を行った場合の正解率は 97.9%となり、高い性能を示した。

発表[2]では、低解像度ナンバープレート数字認識において、PCA (principal component analysis、主成分分析)、スパース PCA、ICA (Independent Component Analysis、独立成分分析) による固有ベクトルを特徴量とし、ベイズ識別してその結果を比較した。「3」は ICA が

最も高い確率で認識した一方で、「9」はすべての手法で認識を誤るなど、不安定な識別結果を示した。今後の課題として SparsePCA のパラメータ調整などが挙げられた。

発表[6]では、日本語情景文字認識の精度を高めるための、複数言語データを用いた Transformer での事前学習手法を提案した。大規模な英語と小規模な日本語を合わせた二か国語データセットでエンコーダを、小規模な日本語データセットのみでデコーダを事前学習することで、エンコーダの頑健性を向上し、デコーダを日本語に特化して文字認識精度が向上することを示した。

発表[7]では、日本古典籍くずし字の認識を目的とし、文字部位に着目した分類手法を提案した。左右や上下などの漢字の構成で分類した場合は高い精度で分類できる傾向を、偏や冠など漢字の部首で分類した場合は精度が出ない傾向を確認した。実施手法の課題としてラベル付けの精度や訓練データの数の偏りなどが挙げられた。

発表[23]では、人間が判別困難なナンバープレートの数字の認識において、CNN での学習のために CG で作成した画像について、どのような画像が学習に有効かを確認した。明暗を変えた画像は有効であり、一部分を切り出した画像は有効でないことを明らかにした。更に、人が判読できない数字を最大平均 98.6%の予測値で推測できることを示した。

[A3 : 文字列認識]

文字列認識に関する発表が 2 件あった[8][9]。

発表[8]では、情景文字画像から文字検出と文字認識を同時に行うテキストスポッティングタスクにおいて、文字検出時の代表点を活用して文字分類時の計算を大幅に削減することによって、精度と速度を両立した文脈自由テキストスポッターを提案した。iPhone 実機での推論速度は最大で 0.01 秒程度とモバイル端末でも実用的な速度で動作することを確認した。

発表[9]では、近年の深層学習を用いた文字列認識の多くが数千万枚の人工合成データで学習されていることを問題点として指摘し、数十万枚の実データによる学習でも同等の精度を達成できることを実験で示した。

[A4 : 前処理]

文字認識の前処理に関する発表が 6 件あった[3][4][10][11][24][25]。

発表[3]は、くずし字の中でも自動での切り分けが難しい連綿体に着目し、切り分けた領域が文字として認識可能であれば適切であると仮定し、学習用データセットの作成と文字認識の精度評価を行った。実験の結果、テストデータの認識精度が 70%程度に留まり、今後の課題としてデータの追加や文字種の追加などを挙げた。

発表[4]は、くずし字によって書かれた古典籍からの自動文字抽出手法を提案した。提案手法は

画像処理やクラスタリング手法を活用し、二値化、エッジ抽出、クラスタリング、過分割した文字の合成と過結合した文字の分割等のステップから構成される。

発表[10]は、低解像度テキスト画像の超解像を目的とし、文字の骨格形状情報を考慮したスケルトン損失関数と大域的なテキスト構造を捉えることができるアテンション機構を組み込んだ超解像ネットワークを提案した。比較実験において、既存手法と比べて提案手法は最も高いテキスト認識率を示した。

発表[11]は、くずし字の文字位置検出を目的とし、U-Net を用いて文字の中心を予測する手法を提案した。評価実験の結果、19 例中 16 例において適合率と再現率の両方が 0.7 を超える性能を示した。

発表[24]は、双二次変換を用いて極端に湾曲した情景文字を正規化することで、文字列認識の精度が 74.0%から 85.1%に向上することを示した。

発表[25]は、くずし字の文字切り出し手法を提案した。具体的には、画像処理により文字領域を抽出し、クラスタリング手法を用いて領域内の文字数を推定し、複数文字から構成された領域に対して文字接続部分を推定することで文字領域を抽出する。今後の課題として、文字領域統合手法と接続部分候補の抽出手法の改善が挙げられた。

[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[12][13][26]。

発表[12]では、GAN (Generative Adversarial Network) を用いた手書き文字列画像の生成において、多クラスで条件付けるバッチ正規化を用いることで、文字セットのサイズに対するメモリ効率が向上されることを示した。また、文字がベース線や中央線を超えているか否かを条件付けることで自然な文字列画像が生成でき、生成された画像における文字認識の精度が上がることを確認した。

発表[13]では、くずし字認識に使用されるデータセットにおける文字ごとの出現回数の偏りが認識精度の低下に繋がることを指摘した。出現回数の少ない文字について、データ数の偏りを補うために画像生成の各手法を適用することで、BicycleGAN は文字の形状を保ちながら多様なスタイルの画像を生成できること、StarGAN-v2 はドメイン情報を与えることで特定の作品に存在しない文字画像を生成できることを確認した。

発表[26]では、競技かるたの試合を撮影した映像について、光や影の影響で札の一部が隠れてしまう問題を解決するため、OCR により抽出された文字同士の隣接関係からグラフを作成し包含関係を比較することで、一部が隠れていてもロバストにグラフを推測する手法を提案した。実際の試合を想定したデータを用いた評価により、短時間でロバストに札の認識を行えることを確か

めた。

B. 文書理解

[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 2 件あった[18][19]。

発表[18]では、動画データを対象とした情景文字認識における訓練データの不足を補うために、1 枚のラベル付き画像から文字列をある軌道に沿って移動させることで疑似的に動画データを作成する水増しを提案した。評価用データセットに対する情景文字認識の性能評価を行った結果、F 値が 46.34%から 74.98%に向上することを確認した。

発表[19]では、ロゴ検出タスクにおいて、学習に合成データを用いることがドメインシフトを引き起こし、性能悪化に繋がっていることを指摘した。また、予測のエントロピーを入力として、それが合成データのものと現実のデータのものを識別する識別器を敵対的に学習することで、ドメインシフトに頑健になることを確認した。

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[5][14][15][16][17][20][21][22][27]。

発表[5]では、1 文字単位の文字画像から筆者を識別するタスクにおいて、ニューラルネットワークの層数を増やし、最終的なマップサイズを小さくすることで、文字形状の表現獲得を抑制し、識別精度を高めたとしている。

発表[14]では、GAN を用いて、指定した印象を持つフォントを生成する方法を提案している。印象のベクトル化に word2vec の重み付き和を用いる工夫によって、既存手法よりも現実に近い画像を生成できることを定量的に確認している。

発表[15]では、ポスターなどの画像中の文字をベクトル化し、指定した文字、色、サイズなどに再編集できる手法を提案している。入力画像を背景、色、テキストの位置と内容、フォントの種類といった要素に分解し、それらから画像を再構成するという 2 段階の構成を一気通貫で学習する手法を採用し、ベクトル化を行わない既存の手法に対して良好な画像を再構成できることを確認している。

発表[16]では、機械学習を用いて、フォントの形と印象との関係を明らかにした。SIFT アルゴリズムによりそれぞれのフォントのパーツを特徴量に変換し、DeepSets モデルに入力してフォントのどのパーツがどんな印象と関係があるかを学習した。フォントのパーツの形や特徴によってフォントの印象が変わることを示した。

発表[17]では、書写文字画像から熟練度を推定する手法を提案した。毛筆書写文字画像に心理

評価値が付与されたデータセットを用いて、畳み込み自己符号化器を用いて文字の特徴抽出をした後、回帰を行った。また、余白除去及びアスペクト比非保存の文字サイズ正規化を前処理として行ったところ、共に精度向上に貢献したことを示した。

発表[20]では、既存の筆跡認証システムの問題点として、筆跡登録者以外の筆跡を考慮されていないという点を指摘した。正解データに本人筆跡を、不正解データに複数人の他人筆跡を用いて2値分類を行う手法を新たに提案し、有効性を示した。

発表[21]では、日本の新大学入試の試行試験において、手書き記述式回答を読み取り自動採点する **end-to-end** のモデルを提案した。文字認識モデルで手書きの回答を認識し、BERT を用いた自動採点モデルに入力した。自動採点モデルの結果は、人間の審査員とほぼ遜色ないレベルであることを示した。

発表[22]では、2つの木簡の破片が同じ木簡からとれたものかどうか判定する手法を提案した。CNN を用いて画像を特徴量に変換し、アテンションベースの VLAD モデルを用いて特徴量をさらに変換する。2つの画像から得られた特徴量の類似度を求めて判定を行うことで、既存のモデルよりも精度良く判定できることを示した。

発表[27]では、既存の文字認識ソフトウェアを用いてナンバープレート認識を行うことを試みた。一般文書と同様に文字認識システムをそのまま適用すると上手くいかないことを確認した。ナンバープレートは一般文書とは文字の配置が異なるため、文字の配置に関する処理を行う必要性があると結論付けた。

表 3.2-1 論文リスト

文献 番号	掲載 媒体	巻号	著者	題目	分類
1	FIT	H-015	楡木徹・浅井紀久夫	CNN を用いた手書き楽譜の音楽記号認識	A2_識別／学習
2	FIT	I-007	大江凌太郎・藤田和弘・四宮康治	SparsePCA を用いた低解像度ナンバープレート数字の識別	A2_識別／学習
3	FIT	F-029	村井 健・市川周	機械学習によるくずし字の文字切り分けに向けた検討	A4_前処理
4	FIT	N-020	高瀬・呂氷・王志辰・孟林	画像処理とクラスターを用いたくずし字セグメンテーション	A4_前処理
5	FIT	H-014	上松和史・荒井秀一	文字の形状情報の学習を抑制することで個人性を強調する CNN 単漢字筆者識別	C0_その他
6	MIRU	I31-29	折橋翔太, 山崎善啓, 牧島直輝, 庵愛, 高島瑛彦, 田中智大, 増村亮	日本語情景文字認識のためのマルチリンガル事前学習	A2_識別／学習
7	MIRU	I31-35	石川太繰, 宮崎智, 菅谷至寛, 大町真一郎	文字部位に着目した古典籍くずし字認識に関する基礎検討	A2_識別／学習
8	MIRU	I31-05	吉橋亮太, 田中智大, 土井賢治, 藤野拓海, 山下直晃	文脈自由テキストスポッターによるリアルタイム・モバイルでの情景文字検出・認識	A3_文字列認識
9	MIRU	I31-33	Jeonghun Baek, Yusuke Matsui, Kiyoharu Aizawa	Scene Text Recognition Only With Fewer Real Labels	A3_文字列認識
10	MIRU	S2-5	中畝志門, 飯塚里志, 福井和広	文字構造を考慮した深層学習によるテキスト画像の超解像	A4_前処理
11	MIRU	I32-01	仲居優志, 宮崎肇, 井上聡, 大山航	深層学習による古文書中の文字位置検出	A4_前処理
12	MIRU	S5-4	Jan Zdenek, Hideki Nakayama	Efficient Model for Handwritten Text Generation with Text Line Awareness	A6_その他
13	MIRU	I32-19	小島朋佳, 山本啓斗, 出川侑暉, 植木一也	くずし字データセットの充実に向けたくずし字画像の生成	A6_その他
14	MIRU	S3-2	Seiya Matsuda, Akisato Kimura, Seiichi Uchida	Generating Font Images with Specific Impressions	C0_その他
15	MIRU	I22-27	Wataru Shimoda, Daichi Haraguchi, Seiidhi Uchida, Kota Yamaguchi	Making Text Editable Again	C0_その他
16	MIRU	I31-02	Masaya Ueda, Akisato Kimura, Seiichi Uchida	Part-based Analysis to Understand Font Impression	C0_その他
17	MIRU	I32-06	吉田直哉, 大山航	畳み込み自己符号化器を用いた毛筆書写熟練度の定量評価	C0_その他
18	PRMU	vol. 121, no. 192	Zhiyuan Xie・Hideaki Goto・Takuo Suganuma	Moving Scene Text Detection Using Synthetic Scene Text Video for Training	B3_文字を含むカメラ撮像画像解析

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	分類
19	PRMU	vol. 121, no. 304	Rahul Kumar Jain ・ Takahiro Sato ・ Taro Watasue ・ Tomohiro Nakagawa ・ Yutaro Iwamoto ・ Xiang Ruan ・ Yen-Wei Chen	Unsupervised Logo Detection Using Adversarial Learning from Synthetic to Real Images	B3_文字を 含むカメ ラ撮像画 像解析
20	PRMU	vol. 121, no. 192	角 大智・久保田 彰	多種の名詞から学習を行った 2 値 分類型の CNN による筆跡認証	C0_その他
21	PRMU	vol. 121, no. 304	Hung Tuan Nguyen ・ Cuong Tuan Nguyen ・ Haruki Oka ・ Tsunenori Ishioka ・ Masaki Nakagawa	Fully automatic scoring of handwritten descriptive answers in Japanese language tests	C0_その他
22	PRMU	vol. 121, no. 304	Trung Tan Ngo・Hung Tuan Nguyen・Masaki Nakagawa	Neural Network-based Local Feature Descriptors for Matching Excavated Mokkan Fragments of Various Sizes	C0_その他
23	情報処 理学会 論文誌	62 (2)	鈴木友哉、宇田隆哉	CNN を用いた予測に有効なナン バープレート写真用トレーニング データの検討	A2_識別/ 学習
24	信学総 大	D-12-14	長岡 巧・若原 徹	双二次変換を用いた極端に湾曲し た文字列の認識	A4_前処理
25	信学総 大	D-12-15	一色昂祐・村木祐太・小 堀研一	くずし字の翻刻のための文字切り 出し	A4_前処理
26	信学総 大	D-20-10	疋田智也・中村開耶・木 村昌臣	グラフ構造を活用したロバストな 札認識手法の提案	A6_その他
27	信学総 大	D-4-10	竹川尚志・金子邦彦	人工知能によるナンバープレート 認識システム設計	C0_その他

4. 文字認識システムの技術調査

4.1 OCR の現状

4.1.1 デジタル化社会時代の OCR の位置付けとその重要性

デジタル庁が誕生し、DX が日本社会を発展させるキーワードとして官民あげて喧伝されるなか、DL を基盤とした AI 技術に注目が集まっている。AI 技術は人との関わりで議論されることが多く、AI 技術の主要な機能としてパターン認識・理解技術の高度化に拍車がかかってきた。そうした時代環境でこれまでデータ入力現場と研究成果との協同経験の豊富な OCR は技術レベルの到達度や未来の発展可能性を探るには大変よい製品である。それは人と人、人と社会をネットワークする一丁目一番地にあるパターン認識技術と、それを応用した入力システムの心臓部となり、あらゆるデジタルシステムの肝心肝要な土台になるからであり、さらに踏み込んで未来を予測すれば、文字読む光学的装置という製品ジャンルから大きく羽ばたき、人類社会の未来に根幹に関わる機械認識の汎用性を高める可能性を秘めたジャンルの製品として実現するという魅力を秘めている。

目標とするデジタル社会実現への道程は AI という強力な武器が欠かせない。デジタル社会を支えるデータは人間社会のあらゆる分野階層から情報を集めそれをあらゆる電算機システムがそれぞれの目的実現のため駆動するが、あらゆる場面で通用する汎用化には AI が欠かせないためである。多様な実効性の高いデータ群を創生するのが入力システムの大きなミッションである。電算機システムは今や全世界をネットワークで結び、あらゆる分野で社会構造を変革しつつあるが、その変革の度合いを決めているのが入力された情報の質である。入力をミスれば自動運転車は人に危害を加えるし、OCR が誤読すればミス補正機能がない OCR システムは社会を混乱させる。機能の魅力を生かすも殺すも入力システムの精度次第であるとも言える。

さらに進んで AI が人間にとって代わるには人間以上に精度が高いというだけでは社会が納得しない可能性が高い。効率の比較からミスもやむを得ないと納得する時代がいずれ来るかもしれないが元も子もなくなるミスは絶対起こさない入力自動システムの実現は恐らくその保証を証明する手段がないと思われる。人がやっても起こす致命的なミスから社会を守るには全く別の安全装置の開発が必須となる。小惑星の自動観測に失敗しても文句は言われないが、核ボタンを誤動作させる入力は許せない。AI を土台に据えたシステムが起こすであろう誤動作の問題は実は近々の課題ともいえる。

4.1.2 OCR の課題と現実的な対応（【参考文献】[1]から[50]）

OCR の課題

先に述べたように OCR の高度化が入力システムの高度化に直結する。ここで現状見えている

高度化の課題についての分析が必要である。（参考文献：本報告書 2 章 3 章など）

それはパターン認識技術の課題でもある。古今東西、未来を予測する機械も人間も存在しない。ニュートンの夢、デカルトの夢、もっと言えば西洋思想の夢といえる科学万能時代は現在の計算機能力ではまだ実現していない。ある意味でもっとも知能が進化したのは人間脳であるが、特に大脳皮質の発達パターン認識をパターン理解のレベルに押し上げたと言えるだろう。進化したのはパターン認識機能である。デジタル革命をリードしてきた電子計算機も近年は加速度を上げて進化しており特異点の誕生を予感させるまでになったが、一方でパターンは無限の細分化を可能とするアナログ的な物理的現実を前提にするならば、特異点の実現はかなり困難な道のりと言える。

例えば、OCR の認識対象として最もポピュラーな帳票を扱う現場の多様性は人間系と物理系の相乗した複雑系であり、予想外、予定外の対象にバリエーションを予測して教え込むのはかなり困難である。というのも現状の OCR の認識技術は過去から積み上がったデータから帳票に記載されたデータを読み取る情報を得ているわけで目の前にある帳票が想定の手帳内にあるという前提で答えを出す技術に立脚しているからであり、過去データから変化の法則性をうまく見つけて近似的に対処する技術開発を積み上げて実用性を高めてきたがまだまだ多くの課題を残しているのが現状である。

OCR の現実的な対応

コロナ（COVID-19）禍はデジタル化遅延の国家レベルでの危機感共有を促しており、その一つの局面として集中する大規模な審査業務における入力効率、コスト問題があぶりだされたといえよう。本来 100%精度の読み取り率であれば OCR 導入ですべてが解決する。情報媒体は紙であれ、オンラインで送られたデジタル映像であれ、人間的なミスを考えなければ一次入力に完遂する。現状では 100%精度の OCR は保証されていない以上、誤読を前提とした入力システムが必要である。元来、入力システムは OCR を使う、使わないにかかわらずミスは機械のみならず人間系でもある確率で起きることを前提で構築されている。DL 等に支えられて AI マシンが急速に発展しつつあるとはいえ現段階では最終的に AI マシンに最終判断を委ねることは車の自動運転と同様に入力システムでも人間系が行う。

つまり、AI-OCR は確かに知識というレベルでの人間系との協調を目指して開発が続いているもののまだ完全に機械任せのシステムの提供はなされていない。したがって、例えば、紙帳票の大量入力審査人間によって業務が完遂する。そのための効率がよく低コストなシステムの提供が求められている [11]。AI-OCR 採用をアピールする入力代行業界でも最終出力は AI-OCR 認識結果に人間によるチェック修正を加えた入力システムによって精度を保証しているのが一般的のようである。

数多くのベンダーが登場し過去にはなかった AI-OCR 百花繚乱状態にある。システムメーカーも自社独自開発以外に最適な他社の認識モジュールを組み込む例もある[37]。

最近の社会的関連では電子帳票保存法改正に対応して AI-OCR を関連付けるところもある。[35、39]。

これからも先進的で多様な社会ニーズに合わせデジタル社会に適応した AI-OCR へと進化を続けていくことが十分に予想される。DL をはじめとした認識理解技術力の相乗的な効果が期待される[38]。

【参考文献】

- [1] DX レポート 2（中間取りまとめ）、経済産業省、2021 年 7 月 5 日差し替え
- [2] DX 戦略、横浜市 HP、2022 年 2 月 3 日
- [3] DXwithSoftBank、SoftBankHP、2022 年 2 月
- [4] DX 推進、[AI-OCR x RPA を活用した DX 戦略、オンラインセミナー、2021.12.107、
https://laqoot.com/](https://laqoot.com/)
- [5] スマホ機能を利用した源泉徴収 OCR 読み取り機能、国税庁 HP、2022 年 1 月
- [6] 源泉徴収 OCR 読み取り機能他リリース、スマート OCR、<https://smartOCR.jp/>、2021 年 12 月
- [7] キヤノン ITS、AI-OCR ソリューション「CapturBrain」無料トライアルプラン提供開始、<http://aiOCR.news>、2021 年 4 月 5 日
- [8] [「スマート OCR 決算書基本パック」を、2021 年 5 月 19 日にリリース、AI-OCR NEWS、
https://aiOCR.news](https://aiOCR.news)、2021 年 5 月 19 日
- [9] エムティーアイ、領収書読み取りアプリ「FEEDER+」の販売、aiOCR.news、2021 年 7 月
- [10] AI-OCR 14 選 解説、起業 LOG、2021 年 12 月 13 日
<https://kigyolog.com/service.php?id=154>
- [11] 紙帳票業務のデジタル化を促進
場所を問わない分散作業を可能にするイメージ審査システムの提供、はいたつく、日立、
2021 年 11 号、p19-p20
- [12] RPA x AI-OCR 導入でテレワーク業務を最低限出社人数で実現、東芝デジタルソリューション㈱、<https://www.toshiba-sol.co.jp/business/rpa/>
- [13] AI×OCR で請求書入力業務を効率化 〜クラウド型 OCR ソリューション | リコー
<https://www.ricoh.co.jp/service/cloud-OCR/special/invoice-system-linkage>
- [14] 一歩進んだ業務自動化・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㈱、
https://www.toshiba-sol.co.jp/pfsol/mojigazou/moji/index_j.htm

- [15] AI-OCR による受注業務効率化ソリューションクラウド型 FAX OCR システム MMLFOS、AI-OCR で誤読率 50%削減、24 時間エキスパートサポート、クラウド型
<https://www.mdsol.co.jp/lp/melfos/>
- [16] 業務用 OCR ソフトウェア DynaEye 10、 PFU、導入実績 7700 社以上、準定型帳票も可能 ヒサゴ帳票も可、スキャナー、複合機など電子化ツール→自動登録 RPA→業務システム、AI でフリーピッチ手書き文字 性能向上、本人確認カメラ OCR、運転免許 OCR、マイナンバーOCR、トータルキャプチャサービス、Google クラウドサービス、RPA サービス、<https://www.pfu.fujitsu.com/dynaeye/product/dynaeye10/>
- [17] 時代を先取りする業務改革と AI-OCR、コロナ禍のテレワーク環境 x 働き方改革、大塚商会、<https://www.otsuka-shokai.co.jp/media/library/material/2011/01006.html>
- [18] [LINE CLOVA OCR 無料体験会、2021 年 12 月 21 日、オンライン、](#)
[LINE CLOVA OCR 無料体験会 | イベント | LINE CLOVA 公式サイト](#)
- [19] 手書き文字の検出、OCR を使用して言語を自動識別、Google Cloud Vision AI ランドマークや製品ロゴ自動抽出、Box で OCR にクラウドコンテンツ管理を導入
<https://cloud.google.com/vision>
- [20] AI よみと〜る、AI-OCR サービス、NTT 東日本、AI inside(株)の DX Suite を活用
https://business.ntt-east.co.jp/service/rpa_aiOCR/
- [21] DX Suite | 最高の AI-OCR を、ディー・キュービック株式会社、
導入事例のご紹介、<https://dx-suite.com>
- [22] AI-OCR ソリューションサービス AI Read、アライズイノベーション(株)、
<https://www.mugen-corp.jp/airead/>
- [23] AI Read とは、CNN,LSTM など DL 活用した AI-OCR ソリューション、
<https://airead.ai/>
- [24] AI-OCR とは、メディアドライブ、人工知能のトレンド含め実践的、体系的解説
<https://mediadrive.jp/technologyAI-OCR.html>
- [25] AI-OCR とは？成功のポイントは認識精度を理解し業務全体を見直すこと、キャノン、
<https://canon.jp/business/trend/AI-OCR>
- [26] AI-OCR とは？導入メリット etc、<https://tegaki.ai>
- [27] AI-OCR とは何か？、<https://ip.tonders.app>
- [28] AI-OCR とは？、<https://tmj.jp>
- [29] AI-OCR とは？OCR との違い、<https://keyman.or.jp>
- [30] AI-OCR って何？、<https://ai-products.net>
- [31] AI-OCR とは、活用事例、製品比較、<https://ledge.ai>

- [32] OCR とは？さらに進化した AI-OCR、<https://rpa/technologies.com>
- [33] 「2022 年 1 月改正」電子帳簿保存、<https://robot-hub-media.bizrobo.com>
- [34] OCR の落とし穴を解説、AI-OCR 導入支援コンサルサービス、<https://metro.co.jp>
- [35] AnyFormOCR、性質違う AI-OCR を同時使用 (W-OCR) 、<https://hammock.jp>
- [36] ICDAR2021、2021 年 9 月 BestPaper、シナモン AI
- [37] シナモン AI、PFU、富士ゼロックス、cinnamon.is
- [38] 事前チェックは不要！請求書・領収書・注文書に特化 [高精度 AI-OCR 対応] OCR ソフト、
<https://panasonic.com>
- [39] WinActor 導入事例・自治体編、横浜市、奄美市、winactor.com
- [40] AI-inside、日本生命 AI-OCR、<https://ai-products.net>
- [41] AI-OCR 「帳票認識サービス」、日立、<https://hitachi.co.jp>
- [42] 手書きの紙帳票からのデータ入力、<https://cweb.canon.jp>
- [43] クラウド型 AI-OCR 「AxisRead」、キッセイ、<https://kicnet.jp>
- [44] 手書き AI-OCR ソリューション、<https://cognet.co.jp>
- [45] AI-OCR の導入事例と効果、<https://braingate.co.jp>
- [46] AI-OCR の有効性、導入メリット、<https://comtix.co.jp>
- [47] 手書き文字でも理論上 100%を実現、[他社の成功パターンから学ぶ AI-OCR x RPA、](#)
[オンラインセミナー、2021.12.10、LAQOOT、https://laqoot.com、](#)
- [48] 【2022 年】AI-OCR 比較 14 選、2022 年 1 月 12 日、<https://boxil.jp>
- [49] AI-OCR 比較 12 選、2022 年 2 月 7 日、<https://aspicjapan.org>
- [50] 【最新比較】AI-OCR 比較 17 選、2022 年 1 月 5 日、<https://kigyolog.com>

4.2 製品分類について

2022年2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)	スキャナ						機能			機構部 HxD×W mm Kg	回路部 HxD×W mm Kg	インタ 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	ホッパ 容量	スタック容量 アクセプト	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
HT-4161	日立	約60 (オプション 使用時約 85)	約120		52x74～ 305x420	約300枚	約300枚	約100(2 枚のみ)	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	約351x約331x約 410(おハ・スタ)約 351x約533x約 410(おハ・スタ)使 用時、 約18.5kg	(機構部に含む)	USB3.0/USB2 .0	GUIによる帳 票定義	2012/7	1x約131万円～ 2x約224万円～ 約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	約407x約715x約 485、 約638x約980x約 485(延長トレイ使 用時)、 約43kg	(機構部に含む)	USB3.0/USB2 .0	GUIによる帳 票定義	2015/7	片面で 約241万円～ 両面で 約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	被写界深 度10mm	—	—	あり	カラー読取り	382x372x270 約3.5kg	(機構部に含む)	USB2.0 (HS)	—	2011/4	オプション 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 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	ホッパ 容量	スタックセ アクト	リジエ クト	解像度 (DPT)	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	(機構部に含 む)	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	(機構部に含 む)	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	OP 両面読取	350x600x450 約30Kg	(機構部に含 む)	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 (オプ 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	OP 両面読取	498x700x430 約75Kg	(機構部に含 む)	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	50mm× 3	50mm	200/240/ 300/400	35kg～ 180kg	OP手書・OP活 字4000種	OP	あり	OP 両面読取	1325x740x19 25 420Kg	(機構部に含 む)	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	OP 両面読取	350x600x450 約30Kg	(機構部に含 む)	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 (オプ 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	OP 両面読取	498x700x430 約75Kg	(機構部に含 む)	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 リユーションテ クノロジー	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	お使用のOS が推奨する 環境以上	2017/2	スーパー版 2,000,000 [税別]	
帳票OCR Ver. 8. 30	パナソニックソ リユーションテ クノロジー	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	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 10/8.1	お使用のOS が推奨する 環境以上	2021/7	クラリアン ト版 398,000 [税別]	
帳票認識ライブラリー Ver. 8. 51	パナソニックソ リユーションテ クノロジー	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手書き文字認識オプ ションあり(価格は別途 ご相談)
GAZOMAGIC	日立チャネルソ リユーションズ 株式会社	手書、活字	数字、英字(大文字)、ひ らがな、カタカナ、記 号、漢字(4133種)	活字：9.5～ 24pt(300dpi)		Windows Vista/7/8.1/10 Windows Server 2012 R2/2016	256MB以上推 奨	2001/9	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	日立ソリユ ーションズ・テク ノロジー	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			Windows Vista/7/8.1/10 Windows Server 2012 R2/2016	1GB以上 (3GB推奨)	2018/10	58,000 [税別]	
帳票マスタ LE V4	日立ソリユ ーションズ・テク ノロジー	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	1GB以上 (3GB推奨)	2013/4	38,000 [税別]	

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]	
FormOCR v. 7. 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	2019/12	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) タイミングマーク	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, 手書きフリーピッチ (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/ライブラリなどの提供形態あり
GLOVIA smart for ファーマシー	富士通					Windows 7/8.1/10	8GB以上	2017/09	980,000円～ [税別]	調剤薬局の処方箋を認識します。
FUJITSU AI-OCR「Keyword Capture Client Edition」	富士通フロンテック					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
FAX/OCR SYSTEM 伝匠 V9	リコージャパン	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 7/10 Server2008R2/2012R2	OCRサーバー: 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。自動帳動方向判別機能。自動帳票識別機能。手書き住所辞書、氏名辞書標準添付。帳票定義ツール付き
FAX/OCR SYSTEM 伝匠 V10	リコージャパン	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 Server2012R2/2016 認識結果修正側: IE 11/Chrome	サーバー: 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(クライアント配布不要)。

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)	識速度(字/秒)	OS	メモリ	発売年月	価格(円)	備考
imageWARE Scan Manager DS V1.1	キヤノン	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、チエックライタイ文字、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	JIS OCR手書き文字(数字、英大文字、カタカナ、記号、マーク) 手書き漢字(JIS第一水準、3,238文字) 活字(明朝、ゴシック) QRコード(モデル2)	手書き、明朝、ゴシック	読取文字サイズ 最小サイズ 横5×縦6mm 最大サイズ 横12×縦12mm		サーバ:Windows server 2008/2008 R2/2012/2012 R2 クライアント端末:Windows 7/8.1	サーバ: 4GB以上 クライアント端末: 2GB以上		要問合せ	
TeleForm	Hammock	手書き英字/活字(大文字・小文字)、マーク認識、バーコード認識、手書き日本語認識/活字				Windows 7SP1/8.1/10(32bit/64bit) Windows Server 2008R2/2012 R2	4GB以上 (OS推奨値以上)		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	アイザイノベーション	活字、手書き	ー	-	-	-	-	-	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

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 (メデイアドライブ)	約6700+英語+中韓語 以下オプシオン 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 2022/2019/2016/2012R2/2012) 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]	中韓語認識オプシオンあり
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 14～15 Android 8～12			2012/6	550,000円 [税別]	中韓語認識オプシオンあり 字種： オプシオン認識対象文 中国語(簡体字) 6763 中国語(繁体字) 13053 韓国語Hangul 2350 韓国語Hanja 4888
GAZOMAGIC LIGHT	日立オムロン ターミナルソリューションズ	活字 (日本語、英語、簡体字中国語、繁体字中国語、韓国語)	明朝体、ゴシック体、教科書体	活字5～72pt ※300～600dpi		VxWorks/Windows/Linux	-	50MB以上	2007/11	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 TM , iOS Linux, Symbian, WindowsMobile, etc.)				要問合せ	

表4.2-8 ソフトOCR製品(文書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	TWAIN 対応	メモリ	発売年月	価格(円)	備考
活字認識ライブラリー Ver. 15.6	パナソニック ソリューションズ テクノロ ジー	約6,800字(英字、数字、ひらがな、カタカナ、JIS記号(一部)168字、ギリシャ文字(一部)32字、JIS第1水準漢字、JIS第2水準漢字)	マルチフフォント (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の場合)		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以上推奨)	2021/10	550,000円 [税別]	
活字認識ライブラリー for iOS /Android V15.6	パナソニック ソリューションズ テクノロ ジー	約6,800字(英字、数字、ひらがな、カタカナ、JIS記号(一部)168字、ギリシャ文字(一部)32字、JIS第1水準漢字、JIS第2水準漢字)	マルチフフォント (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の場合)		iOS 11/12/14 Android 9/10/11		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 [税別]	
読取革命Ver.16	ソースネクスト	約6800	マルチフフォント (明朝体、ゴシック体、教科書体、ワープロ体、新聞文字、斜体など)、英語もマルチフフォント (Century、Helvetica、Courierなど)	6～60ポイント (400dpi)		Windows 11/10/8.1	○	お使いのOSが推奨する環境以上	2020/10	11,800円 [税別]	「簡単！PDF変換」 「簡単！PDF for Office」 「クリップボードOCR」 「フオルダーウォッチャー」 同梱
本格読取5	ソースネクスト	約4000	明朝、ゴシック、教科書体、正楷書体、ワープロ書体、新聞文字など、英語はマルチフフォント対応	6～60ポイント (400dpi)		Windows 11/10/8.1 (32bit/64bit)	○	-	2016/11	3,980円 [税別]	「瞬間テキスト」 「おまかせ名刺管理3」 同梱

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 (メディアアドラ イブ)	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 (メディアアドラ イブ)	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 2022/2019/2016/2012R 2/2012) 対応版あり、 OLEオートメーション 開発キットオプシヨン あり
ドキュメント リーダー 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 ケティンング ジャパン	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、M I C R、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 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、チェックマーク、 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 ソリューションズ インテックノロ ジー	約6,800字	マルチフフォント(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	パナソニック ソリューションズ インテックノロ ジー	約6,800字	マルチフフォント(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		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]	
名刺認識ライブラリー Ver. 3.10 for Android	パナソニック ソリューションズ インテックノロ ジー	約6,800字	マルチフフォント(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)	—	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5おまかせ名刺管理3	ソースネクスト	約4000	明朝、各ゴシック、教科書体、正楷書体、丸ゴシック、ワープロ書体	6～60ポイント (400dpi)	—	Windows 11/10/8.1	○	1GB以上	2016/11	4,378(税 込)	
名刺認識ライブラリー v. 8.0	NTTデータNJK (メディアアド イブ)	約3700+英 語+中韓国 語	明朝、ゴシックほかマ ルチフフォント	5～36ポイント		Windows 11/10/8.1 Windows Server2022/2019 /2016/2012R2/20 12			2016/1	500,000 [税別]	名刺レイアウト自動解 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 (メディアアド イブ)	約3700+英 語+中韓国 語	明朝、ゴシックほかマ ルチフフォント	5～36ポイント		CentOS 7 Red Hat Enterprise Linux 6/7/8			2015/9	500,000 [税別]	名刺レイアウト自動解 析
やさしく名刺 フアイリング PR0 v.15.0	NTTデータNJK (メディアアド イブ)	約3700+英 語+中韓国 語	明朝、ゴシックほかマ ルチフフォント	5～36ポイント		Windows 11/10/8.1	○	1GB以上	2017/11	7,800 [税別]	

表4.2-11 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 (メディアド ライズ)	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 (メディアド ライズ)	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 (メディアド ライズ)	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 カード全体がちょうど 収まるサイズの場合、 1.8 ～ 10 mm 角程 度)		iOS 14～15	—		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 (メディアド ライズ)	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 カード全体がちょうど 収まるサイズの場合、 1.8 ～ 10 mm 角程 度)		Android 8～12	—		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 (メディアド ライズ)	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／一	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 (メディアアドラ 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 (Windo ws Server 2022/2019/2016 /2012R2/2012) 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 (メディアアドラ 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]	
マイナンバー カード認識ライ ブラリー V2.0	パナソニック ソリューション システクノロ ジー	約6,800字	—	6～60ポイント (400dpi)		Windows 11/10/8.1 Windows Server 2022/2019/2016/20 12 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]	
マイナンバー カード認識ライ ブラリー for iOS V2.0	パナソニック ソリューション システクノロ 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]	
マイナンバー カード認識ライ ブラリー for Android V2.0	パナソニック ソリューション システクノロ 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]	

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
免許証認識ライ ブラリー Ver3.0	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、 ワープロ体、新聞 文字など)	6〜60ポイント (400dpi)		Windows 10/8.1 Windows Server 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)	2021/10	550,000円 [税抜]	
免許証認識ライ ブラリー Ver3.0 for iOS/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、 ワープロ体、新聞 文字など)	6〜60ポイント (400dpi)		iOS 12/13/14 Android 4.4/5.0/5.1/6.0/7.0	—	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]	
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 ラットベッ 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4.2-14 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
免許証認識OCR パッケージ	日立	数字、英字、ひらがな、 カタカナ、記号(一部)、 JIS第1水準漢字、JIS第 2水準漢字(一部)	オムニフフロント	1,600画素× 1,200画素(推 奨400dpi相当)	—	Windows 10 Pro/ Windows 10 Enterprise/ Windows 8.1 Pro/Windows8.1 Enterprise	—	インテル Core i5 プ ロセッサ (3GHz)以 上、 メモリ2GB以 上推奨	2016/10	要相談	運転免許証のデジ タライズによるカ ラ画像ファイル からの本人確認に 必要な住所・氏名 等の認識をサポー ト(マイナナンバー カード認識可能)
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	
マイナナンバー解 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	オムニフオン 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、 メダイア別 途)	運転免許証、マイナ 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 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 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 (メダイアド 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サーバー: Windows Server 2019/2016/2 012R2/2012 クライアン ト: Windows 11/10/8.1	-	OCRサーバー: 3GB以上 クライアン ト: 2GB以上	2019/5/27	要問合せ	※オプシ ョン 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	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 (メディアードライ ブ)	約3700	明朝、ゴシックな どのマルチフオン ト	5～45ポイント		依存なし(JSON フォーマットイ ンターフエー 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 (メディアードライ 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 体、他マルチフオ 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)		依存なし(JSON フォーマットイ ンターフエー 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 (メディアードライ 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 体、他マルチフオ 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)		依存なし(JSON フォーマットイ ンターフエー 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 (メディアードライ ブ)	約3700+英語+中韓国語	明朝、ゴシックほ かマルチフオン ト	5～36ポイント		依存なし(JSON フォーマットイ ンターフエー 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 (メディアードライ ブ)	約3700+英語+中韓国語	明朝、ゴシックほ かマルチフオン 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 ソリューションズ テクノロジーズ	<ul style="list-style-type: none"> ・英数字 ・英字、数字、記号の一部 ・住所 ・氏名 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)	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) <ul style="list-style-type: none"> ・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)		—	—	—	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 ソリューションズ テクノロジーズ	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
クラウドOCRサービス Ver. 2.40 (免許証)	パナソニック ソリューションズ テクノロジーズ	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
クラウドOCRサービス Ver. 2.40 (領収書/レシート)	パナソニック ソリューションズ テクノロジーズ	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	パナソニック ソリューションズ 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	Standard (AI手書き、 AI活字) 初期費用 200,000円[税別] 月額 100,000円[税別] 無料枠 50,000ファイル ドまで Pro (AI手書き、AI活 字) 初期費用 200,000円[税別] 月額 200,000円[税別] 無料枠 200,000ファイ ールドまで Basic (AI活字) 初期費用 100,000円[税別] 月額 15,000円[税別] 無料枠 100,000ファイ ールドまで	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
RICOH Cloud OCR for 請求書	リコー	漢字、ひらが な、英大文 字、英小文 字、数字、全 角カタカナ、 半角カタカ ナ、記号	—	—		Windows 8.1 /Enterprise/Pro fessional (32/64bit) Windows 10 Enterprise/Prof essional/Educat ion (32/64bit)	512MB以上の 空き容量	2019/1	月額利用料金 15,000円～[税別]	対象帳票：請求書 連携可能な会計ソフト： SMILE V 会計、PCA会計 DX、大蔵大臣NX、勘定奉 行i 10、勘定奉行i 11、 勘定奉行クラウド
RICOH Cloud OCR for 納品書	リコー	漢字、ひらが な、英大文 字、英小文 字、数字、全 角カタカナ、 半角カタカ ナ、記号	—	—		Windows 8.1 /Enterprise/Pro fessional (32/64bit) Windows 10 Enterprise/Prof essional/Educat ion (32/64bit)	512MB以上の 空き容量	2020/2	月額利用料金 15,000円～[税別]	対象原稿：納品書 連携可能な仕入管理ソフ ト：PCA商管DX、SMILE V 販売、蔵奉行i 10、蔵奉 行i 11、蔵奉行クラウ ド、販売大臣NX

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	オムニフント	6～40ポイント	5秒以下/1枚	ー	ー	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
マイナバンカード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	オムニフント	6～40ポイント	4秒以下/1枚	ー	-	2015/12	月額利用料金45万円～ (初期構築別途)	マイナバン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	オムニフント	6～40ポイント	4秒以下/1枚	ー	-	2015/12	月額利用料金45万円～ (初期構築別途)	運転免許証、マイナバンカード、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。
AI Read on Cloud	アライズイノベーションズ	活字、手書き	ー	-	-	-	-	-	要問合せ	事例：注文書、請求書、決算書、アンケート用紙、チェックリスト共通伝票、特許関連書類
スマートOCR	インフオデオ	-	ー	-	-	-	-	-	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 オンラインミスもあり
Flax Scanner	cinnamon	対象言語：日本語、英語	ー	-	-	-	-		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

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
Tegaki	コージェント ポ	手書き	一	-	-	-	-	-	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
CLOVA OCR	LINE CLOVA	-	一	-	-	-	-	-	Template 月額 55,000円(税込) General 月額 55,000円(税込) 特化型 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
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 6,000リクエスト Standard [税別] 初期費用 200,000円 月額 100,000円～ 無料枠 50,000リクエスト Pro [税別] 初期費用 200,000円 月額 200,000円～ 無料枠 200,000リクエスト	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6万まで 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（または指）とタブレット装置（または専用の用紙）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。

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。しかしながら、新しい技術の展開については、限られた少数の製品を除くと既存製品の維持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は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 を発売しており、手書き文字認識機能が標準搭載されている。昨年は 2020 年 8 月に「Galaxy Note20」、
「Galaxy Note20 Ultra」が米国で発表され、国内では「Galaxy Note20 Ultra」が NTT ドコモと au から発売された（2020 年 10 月～11 月）。その後継機種は 2021 年 1 月に米国で発表、発売された「Galaxy S21」シリーズ（S21, S21+, S21 Ultra）である。このうち「Galaxy S21 Ultra」のみ、S ペン対応となっている。また、2022 年 2 月には「Galaxy S22」シリーズも発表・発売され、「Galaxy S22 Ultra」が S ペン対応である。従来、Galaxy シリーズではペン入力機能を中心とした機種は「Galaxy Note」と呼んでいたが、今後 Galaxy Note という名称を使うかどうかは不透明。現在は「Galaxy S22 Ultra」がペン入力の主力製品である。

Galaxy Note の特徴である S ペンは電磁誘導式のペン入力機能であり、専用アプリの S ノート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。Galaxy Note20 Ultra では、S ペンの遅延が 42ms から 9ms へと大幅に短縮されたが、Galaxy S22 Ultra では遅延が更に短縮されたそうである[1、2]。

・ Arrows シリーズ

富士通の Arrows シリーズ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が可能なシームレス入力に加え、2019 年モデル（Arrows Be3）では

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た[3]。2021 年モデルの Arrows We 等にも踏襲されている。

- **Xperia シリーズ**

ソニーのスマートフォン Xperia は、慶応大の増井俊之教授（元ソニー）が開発した予測入力 IF (POBox) をルーツとした「PoBox Touch」を採用していたが、2019 年 10 月発売の Xperia 5 では機種が未対応であるため Google 社製の Gboard を採用している[4]。元々 PoBox は手書き入力を志向して作られたもので、予測候補の選択操作はペン操作との相性が良いと言われている。多くのファンもいるようだが、2020 年版の Xperia でも採用されておらず、公式に PoBox からは撤退したようである[5]。

- **URBANO、DIGNO シリーズ**

京セラの URBANO シリーズ、DIGNO シリーズは、数年前から東芝の枠なし手書き文字認識「LaLaStroke」を採用している[6]。LaLaStroke は枠内に一文字ずつ記入する「枠あり文字認識」と、複数文字を一気に書ける「枠なし文字認識」、さらに、一画ずつ文字を枠内に記入する「重ね書き文字認識」という複数のモードを持つ。（LaLaStroke の説明ページ[7]は 2020 年 10 月 23 日に閉鎖されたが、URBANO への LaLaStroke の提供は継続している）

- **HUAWEI タブレット[8、9]**

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。

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[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 は Windows10 に簡易版が同梱されており、また、各 OS 用のアプリストアから無料でダウンロードできる。マイクロソフトはアプリストア版の使用を推奨している。

- **Neo Studio[12、13、14、15]**

Android、iOS のスマートデバイス用アプリケーション Noe Studio は（2020 年まで Neo Notes）、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

の2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[16、17、18]

MyScript Nebo は、iPad Pro と Apple Pencil 用に最適化されて作られた iPad 用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。2021 年 7 月に最新の Nebo 3.3 がリリースされた。

- Google 手書き入力[19、20]

Google は、日本語入力アプリ「Google 日本語入力」を手書きに特化させて多国語化した Android 用手書き入力アプリ「Google 手書き入力」を 2015 年 4 月 15 日にリリースした。リリース時には日本語（楷書・草書体）の他、82 言語に対応しており、現時点（2020 年 1 月）には 100 言語に対応と記されている。2018 年には Android 版ソフトキーボード「GBoard」が手書き入力のサポートを開始し、手書き認識の最新機能は GBoard にのみ導入されることとなっている。

- その他[21]

サードパーティー製のソフトキーボードとしては、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」[22、23、24]

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（2022 年 2 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

スマホ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[25]。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[26]

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 等に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[27]。

- Wacom AES 方式[28、29]

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）[30、31]。2017 年 8 月に発売された「Surface Pro」(Surface Pro 4 の次機種なので「Pro 5」と呼ばれることもある)では、新「Surface Pen」

が採用された。これは前モデルの 1024 段階の筆圧検知を上回る 4096 段階の筆圧検知に対応し、より滑らかな入力ができる。また、新たに傾き検知にも対応し、ペンの傾きによって線の太さが変わるため、従来品より本物の筆記具に近い書き味が実現されている[32]。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が、新ペンの方が格段に使い勝手が良いと報告されている[33]。

Surface Pen は最新の MPP (Microsoft Pen Protocol) v1.8 に対応しており、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。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仕組み。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できる[34]。

- 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 等では、第一世代 Apple Pencil が使えるようになっているが、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[35]。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[36]。

なお、Apple Pen は iPad Pro 本体とのペン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[37]。これは 2017 年 1 月発売の新型タブレット「Intuos Pro」にも採用されている。Intuos Pro には「Paper Edition」もあり、タブレット上に紙を置いてクリップで固定し、紙に筆記すると同時に筆跡データの取得が可能となっている。そのために、ペン先にゲルインクペンを採用した「Wacom Finetip Pen」と、ボールペンを採用した「Wacom Ballpoint Pen」も同時に発売している。2019 年 2 月にはスリムタイプの「Wacom Pro Pen slim」も発売されている[38]。

- 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 Note 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。2021 年発売の Galaxy Note 22 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が、ペンそのものには大きな機能的な変化は無いようである[39]（5.1.1 節参照）。

5.1.5 その他の製品・サービス

- ・ Evernote の手書きメモ機能

Evernote はドキュメントに特化したクラウドサービスである。2014 年度から手書きメモデータを直接クラウドに送って格納できるサービスを開始した[40、41、42]。Android 用アプリは 2014 年から、iOS 用アプリは 2015 年から手書きに対応している。

- ・ SkyCom 社 SkyPDF Touch Ink for win 7 [43、44]

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[45]。

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[46、47]。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。

2013 年度にはパナソニックが開発した手書き認識エンジン「楽ひら」をベースとした「楽ひら for Education」が発売された。開発元の PUX 株式会社はパナソニックと任天堂の合弁会社である[48]。(PUX 株式会社[49]は画像認識を中心としたソリューションを提供する会社だが、Web ページに「楽ひら」の紹介は存在しない。現時点でサポートが続いているかどうかは不明である。)

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[50、51]

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。

- ・ 東芝

東芝は独自の「重ね書き文字認識技術」を持っている。これは狭い画面上に重ねて書き続けられるもので、原理はサブストロークベースの HMM である[52]。おそらく「LaLaStroke」で使われている技術に相当すると思われる。

- ・ MyScript (Vision Objects) [53]

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。

- ・ 手書き文字認識エンジン Tomoe[54、55]

Tomoe は、オープンソース環境において手書き文字入力を実現するソフトウェア、また、それを開発するプロジェクトの名称である。機械学習アルゴリズム SVM を用いたフリーソフトウェアのオンライン手書き文字認識エンジン Zinnia を使用し、日本語と中国語簡体字を対象とする。Zinnia は、文字のストローク情報を座標の連続として受け取り、スコア付きで上位 N 候補の認識結果文字を返す。認識速度は、50～100 文字／秒程度である。Sourceforge の Tomoe のページは存在しているが、最終 Update が 2012 年であり、公式 HP のリンクも切れているので、開発・メンテナンスは終了しているようである。

- ・ 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[56]。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、富士通ビー・エス・シー（2021 年 4 月より富士通に統合）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[57]。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。

【参考文献】

- [1] サムスン、「Galaxy S21」シリーズを発表--最上位 Ultra は S ペン対応
<https://japan.cnet.com/article/35165076/>
- [2] サムスン「Galaxy S22 Ultra」発表、「S ペン」内蔵--「Note」シリーズ継承？
<https://japan.cnet.com/article/35183353/>
- [3] arrows Be3 誰でもかんたん機能（重ねて書ける手書き入力）
<https://www.fmworld.net/product/phone/f-02l/simple.html>
- [4] Wikipedia “SO-01M” <https://ja.wikipedia.org/wiki/SO-01M>
- [5] ソニーモバイル製アプリに関する情報や重要なお知らせ
<https://www.sonymobile.co.jp/support/app/>
- [6] URBANO PROGRESSO 設定ガイド
<https://www.kyocera.co.jp/prdct/telecom/consumer/lineup/urbanoprogresso/guide/guide09.html>
- [7] 文字認識ソフトウェア LaLaStroke
https://www.toshiba-sol.co.jp/media/moji/index_j.htm
- [8] HUAWEI MatePad T 11 公式ページ
<https://consumer.huawei.com/jp/tablets/matepad-11/>
- [9] HUAWEI M-Pencil（第2世代）
<https://consumer.huawei.com/jp/accessories/m-pencil-2nd-generation/>
- [10] OneNote デジタルのノートブック
<https://www.microsoft.com/ja-jp/microsoft-365/onenote/digital-note-taking-app>
- [11] 無料で使えるマイクロソフトの OneNote とは？ <https://allabout.co.jp/gm/gc/453716/>
- [12] Neo Notes → Neo Studio 移行ページ <https://www.neosmartpen.com/jp/neonotes/>
- [13] Neo Studio <https://www.neosmartpen.com/jp/neo-studio/>
- [14] おすすめスマートペン：Neo Smartpen 用アプリ Neo Notes レビュー（設定編）
https://www.scary-gadget-life.info/entry/Smartpen_02
- [15] おすすめスマートペン：Neo Smartpen 用アプリ Neo Notes レビュー（使用編）
https://www.scary-gadget-life.info/entry/Smartpen_03
- [16] Nebo 公式ページ <https://www.nebo.app/ja/>

- [17] iPad+MyScriptNebo+Apple pencil - tomi_kun's diary
<https://tomi-kun.hatenablog.com/entry/2019/05/24/003517>
 - [18] “Nebo 3.3: smarter note-taking throughout your workflow”
<https://medium.com/@myscriptapps/nebo-3-3-smarter-note-taking-throughout-your-workflow-d5bac7943524>
 - [19] Google、絵文字も手書きで入力できる文字入力アプリ『Google 手書き入力』をリリース！
<http://octoba.net/archives/20150416-android-news-google-handwriting-ime.html>
 - [20] Android 版「Gboard」が手書き入力をサポート
<https://jetstream.bz/archives/69831>
 - [21] mazec 公式ページ <http://mazec.jp/>
 - [22] Google 入力ツール「手書き入力」
<https://www.google.com/intl/ja/inputtools/services/features/handwriting.html>
 - [23] Google 入力ツールで手書き入力
http://www.libtech.co.jp/2015/05/28/google_入力ツールで手書き入力
 - [24] サポートしている言語 <https://www.google.com/intl/ja/inputtools/help/languages.html>
 - [25] お絵かきペンは“Wacom>N-trig>Synaptic”だと思う！—デジタイザーと Windows タブについて想うコト <https://mupon.net/digitizer-protocol-best>
 - [26] ワコムのテクノロジー Electro-magnetic Resonance
<https://www.wacom.com/ja-jp/for-business/technologies/emr>
 - [27] 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
 - [28] ワコムのテクノロジー Active Electrostatic
<https://www.wacom.com/ja-jp/for-business/technologies/aes>
 - [29] Wacom AES 対応デジタイザープンの“最強”はどれだ？
<https://mupon.net/wacom-aes-no1-pen/>
 - [30] 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>
 - [31] 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>
 - [32] 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>
 - [33] [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>

- [34] 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>
- [35] [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/>
- [36] Apple Pencil とロジクルの Crayon、どちらを買うか本気で考えた
<https://news.mynavi.jp/article/20190617-843972/>
- [37] 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>
- [38] 「ペンが走ってお絵描きが楽しい」ワコムの液タブ・ペントブ用のスリムペン「Wacom Pro Pen slim」レビュー <https://gigazine.net/news/20190227-wacom-cintiq-pro-pen-slim/>
- [39] 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>
- [40] iPad・iPhone で手書き入力に対応 <https://blog.evernote.com/jp/2015/11/05/50527>
- [41] Evernote for Android に手書き入力機能が追加 手書きとテキストが混在させられるのが便利 http://goryugo.com/20140308/evernote_android_handwriting/
- [42] Penultimate – デジタル手書きアプリ <https://evernote.com/intl/jp/penultimate/>
- [43] 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/
- [44] 手書きの電子サインでセキュアな証明—スカイコム・ワコム
<https://www.pc-webzine.com/entry/2020/04/post-358.html>
- [45] 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/>
- [46] 手書きの文字が持つ大きな魅力 <http://www.nintendo.co.jp/nom/0803/p1/>
- [47] 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/>
- [48] 楽ひら for Education <http://news.panasonic.com/jp/topics/2013/37958.html>
- [49] PUX 株式会社 <https://www.pux.co.jp/>
- [50] 日本語文字列認識エンジン
https://ilabo.biz/ilabo_japanese_engine/
- [51] 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

- [52] 登内洋次郎、浅野三恵子、太平英貴、CELL レグザ™リモコン用タッチパッドのストローク認識技術、東芝レビューVol.65 No.4 pp.31-34 (2010)
http://www.toshiba.co.jp/tech/review/2010/04/65_04pdf/a09.pdf
- [53] MyScript Labs 手書き認識のリーダー
<https://www.myscript.com/ja/handwriting-recognition>
- [54] Sourceforge Tomoe <https://sourceforge.net/projects/tomoe/>
- [55] Tomoe の認識モデルを元につかえる Zinnia というライブラリ
<http://taku910.github.io/zinnia/index-ja.html>
- [56] ポトス株式会社 <http://pothos.to/>
- [57] Ink シリーズ：富士通ビー・エス・シー
<http://www.fujitsu.com/jp/group/bsc/services/ink/index.html> (URL 確認 2022.2.11)

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（その他に TC11 が主催する関連学会には IGS もある）。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。今回の会議において、ICDAR は毎年開催することに決まったので、次回、2023 年の第 17 回（米国・サンノゼ）[4]の次は 2024 年（ギリシア・アテネ）の開催となる。

- 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、2020 年に第 14 回がオンライン開催された（ホストは中国の武漢）。次回は 2022 年 5 月にフランスでの開催が予定されている（ハイブリッド開催の予定）[5]。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 ICFHR に改称。2020 年には第 17 回がオンライン開催された（ホストはドイツのドルトムント）。次回は 2022 年 12 月にインドで開催の予定[6]。参加者は DAS とほぼ同じ 150～200 名程度。

- IGS（Conference of the International Graphonomics Society）

人間の動作や神経科学、手書き文字の解析やヒューマンインタフェースなど、人の動きや感覚のデジタル情報処理を対象とする国際会議である。隔年開催であり、第 19 回が 2019 年に 멕시코で行われた。第 20 回は 2021 年開催予定が延期となり、2022 年に IGS2021 という名称で開催される（スペイン）。参加者は DAS や ICFHR より少なく、100 名程度である。

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 への参加が有用であった。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。偶数年には DAS も同じ年に行われるので、両者の立ち位置は次第に見直されるものと思われる。

ICFHR は文字の中でも特に手書き文字に特化した会議である。ICDAR や DAS でも手書き文字が対象の発表は扱われるが、より深い議論を求めるのなら ICFHR が適切である。IGS は手書き文字の生成や署名認証など動作解析により注目しており、他の会議とはスコープが異なる。本報告書で扱う内容に関しては、ICDAR、DAS、ICFHR の 3 つの会議の動向がより重要である。

6.2 2021 年度の開催状況

2021 年度は、前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、依然として多くの大規模イベントが中止や延期となっている。しかしながら、2021 年初旬にワクチンが開発され、少なくとも一時的には感染者数が世界中で激減した（その後、再び増加）こともあり、多くの国際会議がオンラインからハイブリッド（オンライン参加と現地参加の併用）開催へと移行し始めた。前述の主要な 3 つの国際会議のうち、2021 年度に開催されたのは ICDAR2021 である[3]。昨年度の DAS2020 と ICFHR2020 はいずれもオンライン開催であったが、ICDAR2021 は現地（フランス・ローザンヌ）参加とオンライン参加を併用したハイブリッド開催であった。IAPR の Newsletter 2022 年 1 月号に ICDAR 2021 の Special Issue が掲載されている[8]。また、HI 学会誌にて ICDAR 2021 の参加報告が掲載され、ハイブリッド開催に関する感想が記されている[9]。

オンライン開催は Zoom や WebEX 等のツールを用いて特定の URL にアクセスして発表を視聴し、オンライン会話やチャットによって議論を行うものである。セッションごとに割り当てられた会議室 URL に対して個別議論のための別部屋（ブレイクアウトルーム等）を設けるなど、バーチャル環境でも議論が深められるよう様々な工夫が実施されている。ハイブリッド開催は同じ会議へオンラインと現地（オンサイト）の両方で参加できる形式だが、それぞれの参加者間のコミュニケーションに限界があり、運営方法に苦勞するケースが多い。ICDAR 2021 の場合は、会議運営（発表や質疑応答）の大半をオンラインシステム上で行い、現地参加者も原則としてオンライン参加とすることにより、円滑な運営を実現していた。

オンライン開催は人の長距離移動が不要となり、コストをかけずに参加することが可能となるなどメリットは大きい。リアルタイムでの議論（聴講の後の質疑応答やフリートーク）が難しいという課題がある。COVID-19 が終息した後、どのような開催形式が主流になるかは分からないが、選択肢が増えたことによるチャンスは活かすべきだと考える。

6.3 国際会議 ICDAR2021

表 6.3-1 ICDAR2021 の開催概要

日程	2021 年 9 月 5 日（日）～9 月 10 日（金）
Pre-Conference	9 月 5 日（日）～9 月 7 日（火）
Main Conference	9 月 8 日（水）～9 月 10 日（金）
開催地	スイス・ローザンヌ
会場	ハイブリッド開催（現地会場は下記） Pre-Conference : Swiss Federal Institute of Technology Main Conference : Beaulieu Lausanne Convention Center
主催	IAPR (International Association of Pattern Recognition)
公式サイト	https://icdar2021.org/

表 6.3-2 前回大会との比較

	前回 (ICDAR2019)	今回 (ICDAR2021)
開催期間	6 日間	6 日間
開催時間／日	約 8 時間半 (9:00～17:30)	約 8 時間半 (9:00～17:30)
Tutorial	5 件（トータル約 18 時間）	2 件（約 6 時間）
招待講演	3 件	4 件
ポスター	3 枠（2 時間×3）	2 枠（1 時間半×2）
発表セッション	10 枠＋Journal Session 2 枠	10 枠（Journal Session は無し）
参加者	約 500 人	573 人（現地 199 人、オンライン 374 人）
投稿数	454（採択 227: oral 87, poster 190）	340（採択 182 : oral 40, poster 142）

ICDAR 2021 の参加者は 573 名（opening session での発表なので概算）、そのうち現地参加者は 199 名、オンライン参加者は 374 名である。論文投稿数はトータル 340 本で、採択数は 182 本（採択率 53.5%）であった。表 6.3-2 に前回（ICDAR 2019）と比較したデータを示す。ハイブリッド開催の場合、スケジュールは現地時間に従って設定されるため、開催期間や時間は前回と変える必要が無い。ただし、Tutorial の時間が短い、Journal Session が無い、ポスター発表の時間が短いなど、コンテンツは前回よりも少なくなっている。投稿論文数も減っており、全体として 2 割～3 割程度の減少が見られる。その原因は不明だが、何らかの形で COVID-19 は影響しているであろう。しかし、オンラインを含む参加者数が前回よりも増えていることから、ハイブリッド開催という形式はコミュニティの幅を広げる切っ掛けにもなり得るものと思われる。

ICDAR では前半 3 日間の Pre-Conference において個別テーマに分かれた Workshop が 12 件と Tutorial が 2 件行われ、後半 3 日間に Main Conference が行われる。Tutorial では各テーマの専門家が約 3 時間の講義（とデモ）を行う。今回は下記の 2 テーマが開催された。

[Tutorial]

- Tutorial (D. Kotecha, F. Shah, N. Vyas, V. Lonhare, A. Shah)
“The NLP Canvas!”
- Tutorial (A. Doucet, A. J-Caurant, J. Michael, J-P. Moreux, G. Mühlberger, M. Weidemann)
“The NewsEye pipeline for digitalizing large collections of historical newspapers”

Workshop は ICDAR の本会議とは別に個別テーマで開催する小さな会議である。一般に本会議よりも査読の敷居は低いが、特定のテーマに特に関心がある参加者が多く、深い議論が期待される（WIADAR は本会議のポスター発表にて実施）。下記のうち下線付きの 3 件は今回が初めての実施であり、それ以外は前回 ICDAR 2019 でも実施された。

[Workshop (Pre-Conference にて実施)]

- **ICDAR 2021 Workshop (WS) on Graphics Recognition (GREC 2021, 14th edition)**
- **WS on Camera-Based Document Analysis and Recognition (CBDAR 2021, 9th edition)**
- **WS on Historical Document Imaging and Processing (HIP 2021, 6th edition)**
- **WS on Arabic and Derived Script Analysis and Recognition (ASAR 2021, 4th edition)**
- **WS on Computational Document Forensics (IWCDF 2021, 3rd edition)**
- **WS on Future of Document Analysis and Recognition (FDAR 2021, 3rd edition)**
- **WS on Machine Learning (WML 2021, 3rd edition)**
- **WS on Open Services and Tools for Document Analysis (ICDAR-OST 2021, 3rd edition)**
- **WS on Computational Paleography (Paleo, 1st edition)**
- **WS on Document Images and Language (DIL 2021, 1st edition)**
- **WS on Document Visual Question Answering (DocVQA 2021, 1st edition)**
- **WS on Graph Representation Learning for Scanned Document Analysis (GLESDO 2021, 1st edition)**

[Workshop (ポスター発表にて実施)]

- **WS on Industrial Applications of Document Analysis and Recognition (WIADAR 2021, 2nd edition)**

本会議（Main Conference）では、招待講演が 4 件と、口頭発表とポスター発表が行われた。口頭発表はテーマ別に分かれた 8 のセッションと、Journal 採録論文を紹介する 2 つの Journal Track セッションで、それぞれ数件ずつの発表が行われた。ポスター発表は 1 時間半のセッションが 2 回行われた。また、それ以外に企業の講演とパネルディスカッションを行う Industrial セッションと、コンテストの結果発表を行う Competition セッションも開催された。

[招待講演（Keynote Speech）]

- Award Keynote 1 (Prof. Masaki Nakagawa)
“Toward automatic recognition and scoring of handwritten descriptive answers”
- Award Keynote 2 (Prof. Mickaël Coustaty)
“Complex Document Analysis and Its Impact”
- Keynote Speech 1 (Dr. Prem Natarajan)
“OCR: A Journey through Advances in the Science, Engineering, and Productization of AI/ML”
- Keynote Speech 2 (Prof. Beáta Megyesi)
“Cracking Ciphers with “AI-in-the-loop”: Transcription and Decryption in a Cross-Disciplinary Field”

[口頭発表]

- Oral Session 1: Journal Track 1
- Oral Session 2: Journal Track 2
- Oral Session 3: Historical Document Analysis 1
- Oral Session 4: Document Analysis Systems
- Oral Session 5: Handwriting Recognition
- Oral Session 6: Scene Text Detection and Recognition
- Oral Session 7: Historical Document Analysis 2
- Oral Session 8: Document Image Processing
- Oral Session 9: NLP for Document Understanding
- Oral Session 10: Graphics, Diagram, and Math Recognition

[その他]

- Industrial Session
- Competition Session

[ポスター発表]

- Poster: Session 1
- Poster: Session 2

先にも述べたように、ICDAR 2021 はハイブリッド開催であるが、原則として参加者は常にオンライン会議システムにアクセスし、発表の聴講も質疑応答もオンラインで行うというルールで運用された。つまり、現地参加の人もオンラインアクセスを維持することにより、オンラインを基本としたハイブリッド開催という形式を採用した。これにより研究発表や質疑などは参加者全員が共有でき、コミュニケーションの隔絶を最小限に留めていた。ただし、現地参加者が **face to face** で会話した内容はオンライン側からは見えないので、ある程度の不便さはあったと思う。

ICDAR 2021 は独自のオンライン会議プラットフォームをアクセスポイントとして用いていた。多くの講演はこのプラットフォーム上の会議スペースで、**Webinar** の形で行われる。プラットフォーム上にはライブビデオとチャット機能が存在しており、発表も質疑もこの上で行うことができる。ただし、ポスターセッションの場合は、同時に大量の発表者がいることから、プラットフォーム上から外部の **Zoom** サイトを呼び出し、発表者を **Breakout Room** へ割り振ることによって多数の同時発表を実現していた。極めてシンプルな構成であり、参加者にとっては分かりやすく、不便さをあまり感じないシステムであった。

【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] ICDAR 2023 Homepage <https://icdar2023.org>
- [5] DAS 2022 Homepage <https://das2022.univ-lr.fr/>
- [6] ICFHR 2022 Homepage <http://icfhr2022.org/>
- [7] IGS 2021 Homepage <https://graphonomics.net/igs2021/>
- [8] IAPR Newsletter Volume 44, Number 1, January 2022
<https://iapr.org/docs/newsletter-2022-01.pdf>
- [9] 田中宏、事例紹介 ―ハイブリッド開催の国際会議 ICDAR2021―、ヒューマンインタフェース学会誌 Vol.24 No.1, pp.18-21 (2022)

7. 今後の展望

認識形入力方式は実世界の多様な情報をデジタル世界に汲み上げるための重要な基盤技術であり、我が国でも急務となっているデジタル化を推進するうえでも重要となる技術として注目されている。一方、その活躍の場が拡大するにつれ、技術が用いられる環境は極めて多様となり、環境条件に起因する様々な外乱の影響を受けることになる。これがシステムの性能を大きく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、認識形入力方式を用いた装置の性能を十分に発揮するための条件を、より正確に定義することが可能になる。本年度当委員会では、2019年度に公表した、実世界環境においてデジタルカメラやカメラ付きデバイスで文章・帳票を撮影、認識する際の照明ムラに関する取り扱いの指針を示すための「実世界環境における OCR の利用ガイドライン 照明ムラ版」(JEITA ITR-4010)の議論に引き続き、より広範にわたる外乱要因を現象・原因の観点から網羅的に整理するための議論を行った。また更に、その構造化を進め開発者やユーザーにとって有用なガイドラインとしていくための議論を行った。来年度以降も引き続き議論を進め、可能な限り早期にガイドラインとしてリリースする計画である。

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、高い性能を持つ装置を社会に浸透させるための体制作りに努め、関連する規格に関しては各規格の利用状況を考慮しつつ、規格見直し等に関する応答性を高めていく。

OCR 装置の一覧表においては、引き続き最新情報の調査を行い、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。

今後、コロナ後の状況も注視しながら、引き続き国内外の最新技術の動向とそれを活用したシステムについてより広く調査を進め、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	2022 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