特集 3

文字・文書の認識と理解

坂井 邦夫
常葉学園浜松大学経営情報学部
ksakai@hamamatsu-u.ac.jp

入江 文平
(株) 東芝 社会ネットワークインフラ社
bunpei.irie@toshiba.co.jp

水谷 博之
東芝ソリューション (株)
mizutani.hiroyuki@toshiba.co.jp

ロボットの目の機能の１つとして見た場合，文字認識にはいま何ができていて，何ができていないかを述べている．最近の文字認識は，郵便物の宛先読み取りのように記載内容が分かっている場合や，帳票やビジネス文書の読み取りのように記載様式に関する知識を事前に持てる場合には，かなりのことを行える．しかし，ディジタルカメラで自由に撮られた画像中の文字を読むような場合には，必要な解像度を得ること，画質を向上させること，画面中で文字のある場所を探すことなど，文字を読む以前の前処理段階に問題を残している．文字を読む能力を備えたロボットを実現するためには，これらの点での研究開発の強化が今後必要である．

文字を読むとは

人は，無意識のうちに，持てる知識を総動員して文字(列)が担う意味を把握しつつ読み進む．自国語であるなら，相当の崩し字や低品質の印刷物でも読みこなす．ただし，この能力は，言葉を理解することと，その言葉を表記する文字を覚えることを，ほぼ同時期に行いながら成長することのできた人のものである．成長したあとで，まったく知らない外国語の文字だけを覚えようとしても，それは至難の業であり，とても筆記体を読みこなすレベルには到達しない．このように，人の文字認識能力が言語知識の援用に依っていることは，ほぼ明らかである．そもそも読むということ自体が，上記のように文字列が担う意味を把握することなのである．

一方，機械で文字を読むとは，その文字のコードを情報システムに入力することである．このための技術は「文字認識」と呼ばれ，1960 年代に研究が開始されている[1)]．当初は英数字記号など字種の少ない簡単な字形の文字を，パターン情報だけで１文字単位に読み取ることに努力が払われた．それは入力を要するデータが，これらの文字で表される金額，数量，品番などに限られていたこと，シミュレーション用のコンピュータの性能が，まだ十分ではなかったこと，などに因っている．

その後，情報システムのベースが日本語に移行するにつれ，日本語を読めるOCR（光学式文字読取装置）への期待が高まり，漢字認識の研究が盛んに行われるようになった．この時代には同形異義文字や類字形文字を識

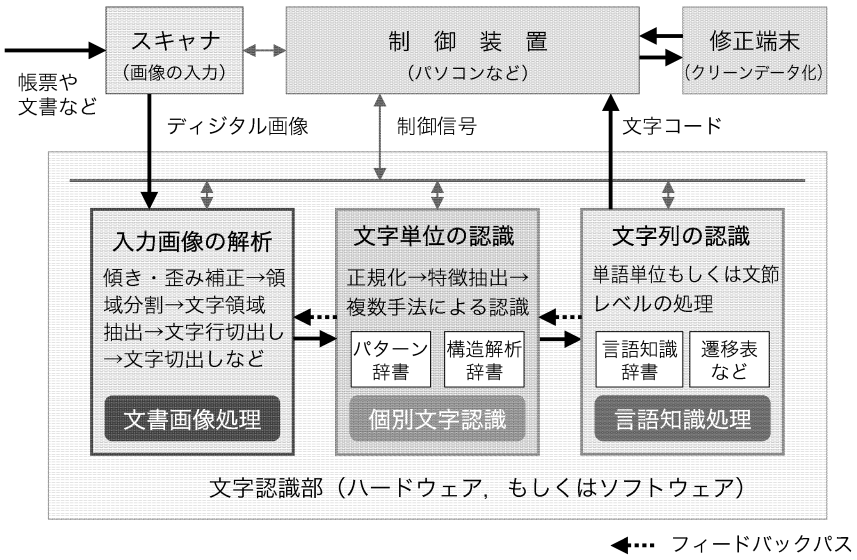

図-1　文字認識システムの構成例

別するため，単語や文節レベルの言語知識処理が導入され，「個々の文字を読む」から「文字列を読む」時代へと進んだ．

さらに，パソコンやワープロの普及がソースデータ（情報システムに入力する元データ）の印刷文書化を促し，これを読み取るOCRに文書のレイアウト構造を解析する機能（文書画像処理）が付加され，「文書全体を認識・理解する」時代になった．この能力，すなわち処理する対象の全体をとらえてその中の文字を読むという能力は，ロボットという新しい情報システムの目の機能の1つとして使えそうである．そこで以下では，文字認識にはいま何ができていて，何ができていないかを，ロボットへの応用を念頭において概観する．

文字を読む技術

前述のように，文字認識は情報システムの発展とともに進化する．ロボットを引き合いに出すまでもなく，情報システムがより人に近いところで使われるようになるにつれ，文字認識には人と同等の機能と性能が求められる．この要請に応えるためには，人の持つ優れた能力を機械にも持たせようとする取り組みが不可欠である．実際，現在の文字認識には，その方法論として，人のパターン認識過程を分析的にとらえて得られた多くの知見が活かされている．たとえば，

- 人は，まず全体を大きくとらえ，その後に細部を段階的・階層的に見て複雑な対象を理解する
  －文書画像処理方式の開発
- 大脳皮質の第1次視覚野には，視覚経験によって組織化される方位選択細胞がある
  －方向性パターンによる特徴抽出の採用
- 論理的な処理と直感的な処理を併せ用いて認識を行う
  －構造解析法と重ね合わせ法☆1の組合せ
- 後天的に獲得した知識を用いて判断を行う
  －言語等による知識処理の導入

などの例がある．

さて，この方法論を具体化するには，画像の入力から文字コードの出力に至る一連の処理をプログラム化し，その過程で参照される何種類もの遷移表☆2や辞書を作ってシステムに埋め込み，全体として動くものに仕立てなければならない．特に重ね合わせ法で用いるパターン辞書の作成は非常に難しく，多くの帳票や文書から少なくとも1字種あたり数千から数万のサンプル文字パターンを収集し，それらの特性をコンピュータで分析する必要がある．分析の結果得られた字種別サンプルパターン集合の縮約表現が標準パターン，これを収録したものがパターン辞書である．辞書にはこのほか，文字パターンの形状を部分輪郭の系列や幾何学的特徴の有無などで記述した構造解析辞書，対象言語の単語と文法を収録した言語知識辞書がある．**図-1**に文字認識システムの構成例を示す．

最近の研究開発

文字認識というテーマは，古くて新しい．その意味するところは，「何か問題が解かれると，次の挑戦課題が実社会から提示され，これに向けた技術開発がまた始まる」ということである．これこそ，40年に及ぶ文字認識

☆1　重ね合わせ法：入力パターンと標準パターンを重ね合わせたときの重なり部分の多少でもって識別・決定を行うパターン認識方式のこと．
☆2　遷移表：特徴種別や文字コードなどの記号系列を受理して動くオートマトンの状態遷移を記述するテーブルのこと．

の研究を支えてきた最大の要因である．実際，昨今の文字認識は，社会的要請に先導されて，「専用のOCR帳票を用意し，入力データの品質を管理し，特定業務のデータエントリを行う」という段階から，「企業や官公庁が現用しているさまざまな帳票（既存帳票）やビジネス文書をそのまま読み取る」という段階へと進化し，新たな用途を開拓している．この方向は，外界にある文字を特定の目的やタスクに縛られず自由に読むというロボットの目の要求仕様にも合致する．

その基になった最近の研究開発は，

- 特徴抽出や識別空間の決定を，大規模な機械学習によって行う辞書設計手法の開発
- 相互に接触もしくは入り組んだ文字の分離・切出しを，個別文字認識や言語知識処理の結果を帰還（フィードバック）させて行う技術の開発
- 帳票書式の自動登録および識別を，文字枠の構造やプレプリント文字の状態をもとに行う技術の開発
- 文字パターンとフォーム☆3パターンの分離を，線特性の違いを利用して行うフォームリムーバル技術の開発
- 多色刷り帳票上の罫線，印鑑，地紋などの検出・分離を，色情報を用いて行う技術の開発
- 文書のレイアウト構造を，文書というものの特性／モデルに基づき解析する技術の開発

などである．

いずれも対象に関する知識を最大限に利用するという点に特徴がある．したがって，上記技術を組み合わせて構成される現在の文字認識は，「知識を総合した文字読み取り」であるといえる．以下にその例を示す．

知識を総合した文字読み取り

郵便物の宛先読み取り

我が国における郵便物区分の自動化は，3～5桁郵便番号制度の制定に合わせて1968年に導入された郵便番号読取区分機に始まる．この装置は赤枠内に一定のルールに従って書かれた手書き郵便番号を読み取るもので，収集した郵便物を差立局において配達局別に仕分けする差立区分の自動化に用いられた．その後，宛名ラベルなどの印刷文字を読み取る機能拡張を経て，1989年には住所読取区分機が導入され，配達局において概ね町名までを読み取り，配達人別に仕分けする作業，つまり配達区分が自動化された．1997年になると，7桁新郵便番号制度が制定され，新たな郵便処理システムが導入された．このシステムでは，住所の丁目番地号，棟室番号等，住所階層のさらに下までを読み取ることができる．

☆3 フォーム：罫線などのプレプリント情報のこと．

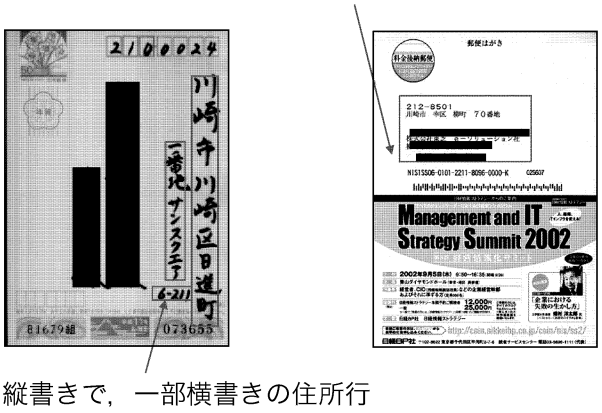

図-2 住所領域／行検出の問題点

宛先読み取りの技術的な特徴は，郵便の宛先というコンテキストのもとで自由に記載された不特定多数者の手書き文字ならびに多種多様な字体の印刷文字を，そのコンテキストに依存する知識を用いて読み取り，宛先コードに直して出力するという点にある．

宛先読み取りの手順は，概ね画像前処理，住所領域／行検出，文字切出し，文字認識，知識処理とに分けられる．

画像前処理では，できる限り鮮明な住所部分の画像を得るため，生のセンサ信号に対し，適応的2値化や空間フィルタリング等の処理が施される．この際には，地模様，セロハン窓の有無など，郵便物の表面状態に関する知識が援用される．

住所領域／行検出では，住所の構造とその記載様式，文字の大きさや並び方に関する知識が用いられる．しかし手書き文字では，縦書き，横書き，縦書きで一部横書きなど，さまざまな変形がある．一方，印刷文字では，広告文や絵などが宛先近くに印刷されると，どこまでが住所領域かの判別が難しくなる（**図-2**参照）．このように住所行を正確に検出することは容易でなく，上記の知識に加えて，後段の文字認識結果が参考にされることもある．

文字切出しでは，文字の大きさや並び方，文字と文字の接触のしかたなどに関する知識が用いられる．通常，この段階では，文字の切れ目の位置候補が複数残され，後段の文字認識や知識処理の結果で確定される場合が多い．

文字認識では，言語依存性を軽減するため（区分機は輸出もされている），機械学習を取り入れた統計的手法が多用される．

知識処理では，地名，住所記載様式などに関する知識のほかに，区分機が稼働する場所や地域に関する情報も使われる．

以上のように，郵便物の宛先読み取りにおいては，パ

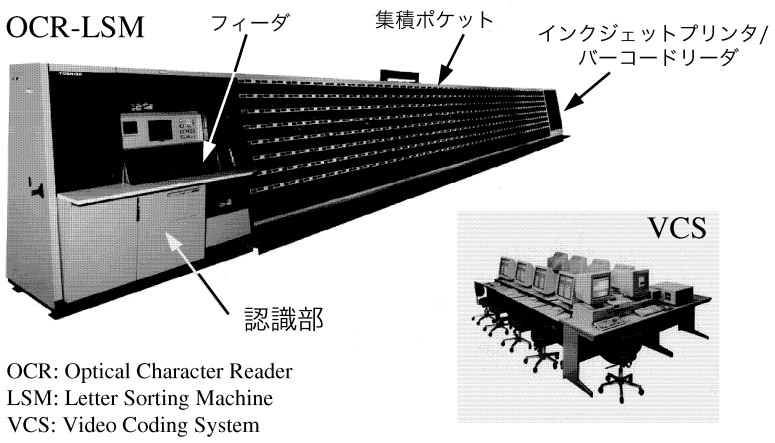

図-3　郵便区分機の概観

図-4　振込依頼書読み取りの問題点

ターン認識や画像処理の技術に加えてさまざまな知識が総動員されている．郵便物を仕分けする専用ロボットとして見た場合，現在の郵便区分機（**図-3**）の処理性能は，VCS（ビデオコーディングシステム）による修正作業を含めて毎時最大３万通である．ここでは，文字単位の読み取り精度よりも，宛先コード単位の正読率の方が重要である．

既存帳票の読み取り

郵便振込みは最も身近な送金手段の１つとして社会に定着し，そのための用紙（振込依頼書と呼ばれる既存帳票）とともに，長年，愛用されている．ただしこのサービスは公共サービスであり，万人が利用するという性格上，用紙はもとより，筆記や印字に関して強いガイドや制限を設けることは難しい．このような利用環境のもとで生成されたデータに対する読み取り性能は，外界にある文字を自由に読むというロボットの目の要求仕様が，現在の技術でどこまでに満たされるかを測る尺度になる．

さて，振込依頼書では，**図-4**のようなケースが入力データにしばしば混入する．振込依頼書OCRは，このような場合であっても，金額や口座番号などの重要データを正確に読み取らねばならない．そのためには，フォームリムーバルの技術が決め手となる．これはドロップアウトカラー[☆4]ではない色で印刷されたフォームのパターンを入力画像から除去し，次に行う文字切出しを可能にする技術のことである．

フォームリムーバブルは一見，やさしいようにも思えるが，そこには簡単な処理では済ますことのできない難しい問題が存在する．たとえば図-4上段のように，記入文字や印刷文字が罫線と交錯している場合には，単純な罫線除去を行うと，文字線の一部が欠落する．これを避けるためには，文字線部分は残して罫線だけを除去するという高度なテクニックが必要である．

現在，郵便局で稼働している振込依頼書OCRには，フォームリムーバブルのほか，筆記具や用紙に起因するさまざまな問題に対処する技術が開発・搭載されてい

☆4　ドロップアウトカラー：OCRのスキャナ部のセンサには感じられない特殊な波長の色のこと．

る．これにより，振込依頼書OCRの帳票通過率（帳票内のすべての文字を正しく読めた帳票の割合）は95%を超えている．つまり，郵便振込み業務のソースデータ入力は，そのほとんどが自動化されている．

既存帳票読み取りの他の典型例は，税務処理である．たとえば，多くの自治体では，中央省庁から回付される税務関係書類や，法人などから収集した納税申告書を処理している．政令指定都市などでは，年間数十万～数百万件におよぶレコードを扱う場合もある．いずれも，ほとんどが既存帳票上に印字された文字データである．

税務処理用のデータエントリは，課税・納税という業務の性質上，要求される精度が高い上に，レイアウトのゆれ☆5，読み取り対象に近接した罫線やプレプリント文字，低品質印字などの技術的問題を抱えている．

そこで(1) レイアウトのゆれに対しては，読み取り領域周辺のパターンを補助情報に用いたマッチング処理により記載欄の正確な位置決めを行う，(2) 読み取りに邪魔となる罫線やプレプリント文字に対しては，これらのパターンの色彩情報を利用した除去を行う，(3) 低品質印字に対しては，雑音生成モデルと機械学習を組み合わせた辞書設計方式を用いる，などの解決策がとられている．

このような対策を講じた結果，現在の税務処理OCRの性能は，既存帳票上の印刷文字認識において，キー入力に肩を並べるところまできている．一例を挙げると，表計算ソフトなどで作成された細かな数字データの読み取りでは，速度はもちろん，認識精度においても数百万文字中にエラー数十文字という，人のキー入力精度（数百万文字中にエラー百文字程度）を凌ぐ結果が得られている．

既存帳票読み取り技術の基本は，罫線やプレプリント文字というゆらぎを伴う背景雑音の中から，認識対象となる文字のパターンをいかに正確に分離して取り出すかというところにある．入力システムとして見た場合の評価基準は，文字単位の誤読率の低さである．

ビジネス文書，書籍・雑誌，新聞などの読み取り

紙というメディアを介して表記される情報の大部分は，ビジネス文書，書籍・雑誌，新聞などのかたちで流通している．したがって，これらの印刷物が担う文字情報を取り込むことは，OCRのみならず，ロボットにとっても必要なはずである．しかし，その多様性のゆえに，これらの印刷物に対して統一的な処理を行うことはほとんど不可能である．文字を読もうにも，それがどこにあるかはまったく不定であり，文字を探すこと自体が難しい．

具体的にいうと，上記印刷物の紙面には，文字（テキスト）以外に図表，写真，画像など，さまざまな種類の文書構成要素が，それぞれある範囲を領有するかたちで混在している．各構成要素の配置は自由であり，またページごとに変化するので，紙面のレイアウトに関する事前知識は存在しない（既存帳票の読み取りとはここが異なる）．知識として使えるものは，縦書き／横書き，章・節・段落，表題，見出し，脚注，ページ番号などテキスト領域に関するおおよその基準，および各構成要素を特徴付ける属性に関する知識☆6である．

このような紙面を走査して得られた文書画像中の文字を読むには，画像のどこにテキスト領域があるかを探すことと，テキスト領域が複数ある場合には，それらの関係（たとえば読み順）を明らかにすることが，先行して必要になる．

レイアウト解析は，これを行うための基本技術である．ここでは，上記の一般的ともいえる知識をもとに，ボトムアップとトップダウンの双方向から入力画像の解析を行う[2]．すなわち前者では，属性に関する知識をもとに，画素という最もプリミティブなレベルから始めて，文字パターンという同一属性を持つ部分を徐々に統合していく（群化）．一方，後者では，文字行などテキスト領域の構成要素をモデル化したルールにより前者の処理を検証し，誤った群化に対しては修正を施す．

業務用のドキュメントリーダ（文書読み取りソフトウェア）では，この方法で，98%前後の成功率で文字行検出を行えるところまできている[2]．

図-5は，新聞紙面（左の図）に対してレイアウト解析を行った結果（右の図）の例である．テキスト領域については，文字行検出までを行っている．

以上のように，ビジネス文書，書籍・雑誌，新聞などからなる多種多様な文書の読み取りにおいては，ボトムアップ処理とトップダウン処理の相互作用を繰り返し，結果として入力画像に適応したテキスト領域等の探索を自律的に行う技術，すなわちレイアウト解析が不可欠である．

モバイル環境での文字読み取り
―ディジタルカメラ入力画像の読み取り―

ブロードバンドネットワークの進展と携帯電話やPDAの普及により，モバイル環境での仕事や生活が可能になりつつある．特に最近では携帯電話にもディジ

☆5 レイアウトのゆれ：記載内容は同じでも，発行元の違いなどにより，帳票上に印刷された記載欄の位置や相対的配置が微妙に異なること．その程度には，かなりの幅がある．

☆6 属性に関する知識：たとえばテキスト領域には横長もしくは縦長の帯（行）があり，かつその中には比較的規則正しく並んだ方形のかたまり（文字）がある，図表領域には長い直線や曲線がある，写真領域には濃度分布における中間調が存在するなどのこと．

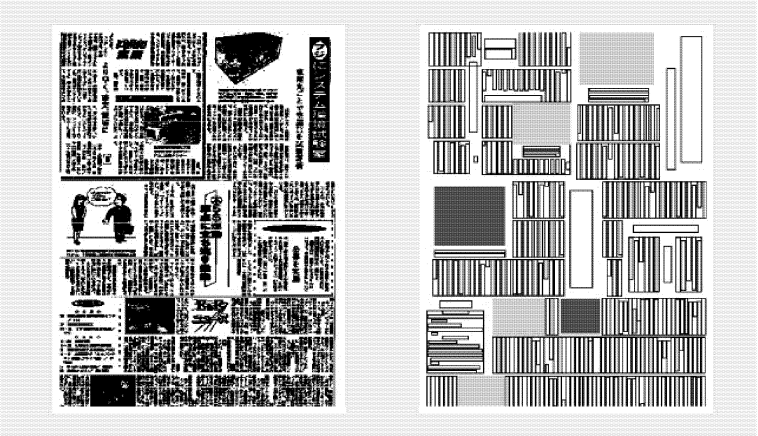

図-5　レイアウト解析の例

図-6　ディジタルカメラ入力画像の例

タルカメラが搭載され，音声のほかに画像（や映像）を含めたコミュニケーションができるようになった．しかし画像の中には，e-mailアドレスやURL，電話番号などのように，文字コードとして扱うべき対象が多く含まれる．このような場合に，画像から文字コードへの変換を行うものが，ディジタルカメラ入力画像の読み取りである．

おそらく多くのロボットの目は，ディジタルカメラに類するものを備えるであろうから，ディジタルカメラ入力画像の読み取りは，「文字を読むロボットの目」的な機能の実現例であるといえる．

ディジタルカメラで入力された画像は，**図-6**のように，従来のOCRスキャナによる画像に比べて画質が大幅に劣る．その主要因は，(1) 撮影条件に起因するもの：照明の具合やフラッシュの焚き方，撮影者の手ぶれや撮影対象に対するカメラの傾きなど，(2) カメラ自身に起因するもの：レンズにおける歪曲収差や周辺部分のピントずれ，受光素子数や画素数の不足などである．これらはそれぞれ，入力画像中の輝度の勾配や局所的な明暗，文字の形状の歪・ぼけ・かすれ・つぶれなどを引き起こす．

正確な読み取りを行うためには，まずは劣化した画像に対する補正前処理，たとえば低解像度画像に対する階層的2値化や変換行列を用いた透視歪みの補正などが必要である．次には読み取りの対象となる文字が含まれる場所，すなわち入力画像中の処理領域を特定する技術が必要である．ディジタルカメラによる撮影対象は，文書，看板，名刺，パネル，帳票など多岐にわたるため，多くの場合，入力画像の中から上記領域を探し出す汎用的なアルゴリズムを開発することは大変難しい．撮影対象に共通して利用できる知識がほとんどないからである．したがって，「画像中央に写っている領域には，読み取り対象が高い確率で含まれる．したがってそこを…」というような簡単な指針に頼ることになる．しかし，複雑な背景に埋もれた未知の入力画像から文字のパターンを見つけること自体が，文字スポッティング[3]という名で呼ばれる難問である．これは今後の課題である．

また，屋外にあって外気にさらされている文字は，紙の上の文字とは異なる変形を受けている場合が多い．このような文字を読み取るためには，個別文字認識のさらなる頑健性向上が必要である．変形に強い特徴量を探すなど，文字認識の原点に立ち戻った取り組みも必要であ

3次元空間で	自律的・能動的に文字を見つけて読む	ロボットの目の機能の1つ	文字らしさの尺度の定量化 →文字スポッティング技術の開発 劣化した画像の修復 →補正前処理技術の高度化 個別文字認識の頑健性向上 →変形に強い特徴量の発見など
	与えられた画像の中から文字を見つけて読む	モバイル環境での文字読み取り ・現状は携帯電話のディジタルカメラ入力画像を読み取るレベル	
2次元平面で	文書全体を認識・理解する	印刷文字 ・既存帳票の読み取り ・ビジネス文書，書籍・雑誌，新聞などの読み取り	レイアウト解析の精緻化には，文章の意味・内容にまで立ち入る必要がある
	文字列を読む	手書き文字 ・郵便物の宛先読み取りなど 印刷文字 ・ワープロ原稿の読み取りなど	自由記載欄に記入された手書きメモを読み取る研究も今後は必要
	個々の文字を読む	手書き文字 ・数字については，誤読ゼロに近い 印刷文字 ・漢字まで高精度に読み取れる	低品質の手書き漢字・英数字カナについては，ブレイクスルーが必要
何をするのか		できていること	できていないこと（課題と対応）

（■：完成度が高い，□：完成度が低い）

表-1　文字認識にできていること，いないこと

ろう．

このように，ディジタルカメラ入力画像の読み取りには，文字認識の進化につながる多くの技術課題が含まれている．この課題への取り組みは，単に携帯電話の機能拡張といったことではなく，これまで紙という2次元平面の上に閉じて使われてきた文字認識の技術を，3次元空間の中でも使えるようにするという意味を持つ．すなわち，次に述べる「ロボットの目の機能としての文字認識」に実現の手だてを与えるものである．

ロボットの目の機能としての文字認識

外界からの情報の約8割は視覚から入るという．そのまた何割かの情報（特にセマンティックなもの）は，文字によって表されたものではないか．運動能力や形をどのように人に近づけたところで，文字も読めないロボットと人との関係は深まりそうにない．少なくとも外界の情報に基づく知的活動の場における共存・協調は望み薄である．この意味で，文字を読む能力を持ったロボットの研究は，全人的ロボットはいうに及ばず，ロボット全体の将来的発展にとって重要であろう．

著者のイメージする文字を読むことのできるロボットの活用シーンは，たとえば，(1) 新聞の見出しを読んで聞かせる―日常生活の場でのサービス，(2) 標識を見て巡回する―警備の代行，(3) 書類や器材を指定した番号の部屋に届ける―オフィス，病院などでの使役，(4) 危険な瓦礫の中から文字を探してその映像を送る―災害救助・復旧の補助，(5) パネル画像を見つけて撮影し，文字部分を読み取る―展示会場での情報収集，(6) 交通標識や店の看板を読みあげる―視覚障害者の外出支援などである．

これまでに述べたように，文字認識の技術は，「ここに文字がある」ということを教えられれば，もしくは暗黙裏にそれを仮定できる場合には，**表-1** のまとめのように，相当のことを可能にしている．しかし，時間的，空間的にさまざまに変化する環境の中で，自律的・能動的に文字を見つけて読み取る機能（前述の文字スポッティング）は魅力的ではあるが，まだその研究は始まったばかりである．ところが人にとってこれは容易なことである．たとえば，外国でまったく知らない文字に出会っても，それが文字であるらしいことの判断はできる．文字という人工パターンと，それ以外のパターンを区別する何らかの尺度を人は持っているに違いない．この尺度が分かって定量化されたとき，「知識を総合し，モバイル環境で文字を読む」ロボットの目は，その利用範囲を大きく拡げることになるであろう．今後の発展に期待したい．

参考文献
1) 中野康明: 文字認識・文書理解の最新動向，電子情報通信学会誌，Vol.83, No.1-No.7 (2000).
2) 石谷康人: データ駆動型処理と概念駆動型処理の相互作用による文書画像レイアウト解析，情報処理学会論文誌，Vol.42, No.11, pp.2711-2723 (Nov. 2001).
3) 小川英光（編著）：パターン認識・理解の新たな展開－挑戦すべき課題，電子情報通信学会 (1994).

（平成15年9月22日受付）