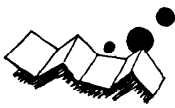


## 解説



# パターン認識技術の実用化†

森 俊 二\*\*

## 1. ま え が き

最近のパターン認識技術の実用化には目を見張るものがある。雌伏三十余年、今や一斉に開花し始めたという観を呈している。この理由は、一言で言えば、'認識アルゴリズムが LSI に乗るようになったからである'ということになる。

しかしながら、ここで一体実用化とは、アルゴリズムとは、LSI 技術とは何んであるのかが問われなければならない。これらの諸問題を考察することによりパターン認識技術の歴史的発展、すなわちその過去と現在、そしてそれを基にしてその未来をかき見ることができよう。われわれが真に知りたいのは、この情報処理技術の中において、本質的と思われながら従来の科学技術とは何か異質であるこの技術の発展のメカニズムである。しかしこの考察は必然的に技術一般の発展のメカニズム、そして技術と社会のかかわり合いという巨大な問題に立ち向わせる。これは勿論、一技術研究者の能力を越えるものである。あえて、しかしこのようなことを持ち出すのは、上述の問題は基本的には、上に述べた枠組みの中で考察されなければならないということを喚起したかったからである。以下上述した諸問題について私見を述べ、可能な限りこの基本的問題について触れて見たい。

## 2. 実用化の量と質

実用化と言ってもその意味するものは広く深い。まずここではその量的側面と質的側面に分けて考察することにする。

### 2.1 実用化の量

実用化の量的側面、すなわちその実用化の形態、またはその規模は大きく三つに分けることができよう。すなわち第一は特殊ユーザへの実用化である。第二は

一般事業体ユーザへの実用化、第三は個人ユーザへの実用化である。第一の典型例が、軍事利用である。好むと好まざるにかかわらず、軍事目的のための巨大な開発研究が技術発展のけん引車であった。情報処理技術においても例外ではない。計算機それ自身がそうであることは衆知の事実である。ともかく大事なことは技術が実用化される過程の中で、この段階は非常に重要であり、第二、第三への発展のための準備期としての役割を果たすということである。特にパターン認識の場合、最初から完全なものではできない。それゆえ'理解あるユーザ'の存在がエッセンシャルである。

非軍事国日本で、パターン認識技術において、第一の役割を果たしたのは、言うまでもなく郵政省による郵便番号制度とその読み取り装置の導入である。これについては玄地氏による詳しい解説が本小特集号にあるのでそれを参照されたい。ともかく、これが世界の最高水準にあるといわれる日本の OCR 技術の基盤を作ったと言って良いであろう。そして現在手書き英数字、片仮名を読み取る OCR は数百カ所の事業所で使用され、台数にして 3,000/年と言われており、第二の実用段階にはいった。また一方印字された英字/片仮名文字については POS 用としてハンディスキャナが出回る気運にあり、第三の実用段階に一歩近づいていると言えよう。米国では、むしろこのハンディスキャナが広く使用されている。

さて他の分野ではどうかと言え、画像処理技術では、その強調、回復技術が人工衛星からの伝送写真の復元を目的として、ジェットプロパルジョン研究所を中心とし米国において大きな発達を遂げた。これも軍事・科学を目標とした政府をユーザとする第一の実用化の典型例である。他の例としてはモートセンシング用に GE によって開発された Image 100 などが上げられよう。このようにいわゆる画像処理技術は第一の実用期にあるが、しかし特に米国で、いくらかでも第二のレベルに近づけようという努力が見られる。著名なこの分野の研究者であった、Andrews と Pratt があ

† Pattern Recognition Technology being Put to Practical Use by Shunji MORI (Computer Science Division, Electrotechnical Laboratory)

\*\* 電子技術総合研究所パターン情報部

い次いで画像表示処理装置専門メーカーに移った話は有名である。

一方ロボット技術はどうかと言えば、広い意味のロボットは特に日本で非常に良く使用されていて、統計<sup>1)</sup>によれば 47,000 ものロボットが働いているということである。しかし米国では一桁以下の 3,255 で、ここに如実に日米の生産性の差が現れているというわけである。今や米国では、生産性向上を旗印としてロボットブームが興りつつある。ともあれ、パターン認識技術の立場からすれば、ロボットは形状を認識できなければならず、現在それができて現場で働いているロボットはそんなに数が多いわけではない。この代表的例が日立の掃除器の自動組み立てシステムであり、本小特集で鳥野氏により詳しく述べられている。米国で代表例を挙げるとすれば GM の Sight-I であろう。これは IC チップの検査システムであり、この種の技術が米国で生産ラインに実際に利用された最初 (1977 年) のものであると GM では言っている。この詳しいこととその他の例についてはロボット特集号である文献 2) を参照されたい。またロボットではないが形状認識技術の実用化成功例として良く知られているものに三菱のきゅうりの自動選別システムがある。また手書き OCR 技術の自然な延長として手書き図面の入力装置が強く望まれていたが、それを初めて実用化したのが富士通の手書き図面の自動入力装置である。これは社内で実際に良く利用されているということである。これらについて本小特集号においてそれぞれ前田、伊藤の両氏、吉田氏による紹介がある。この種の技術利用の潜在性は膨大で、もしそれが実現すれば社会問題を引き起こしかねないと思われるほどである。しかしこれは今のところは杞憂にすぎず、以上述べたごとく特殊用途に限定されていて、第一の実用段階に留まっている。

さて音声認識技術の分野ではことの本質上、バッチ大量処理といかないからか政府のような強力なユーザーの利用というものはないようである。むしろ反対に最近この技術に近い音声合成技術が、玩具とはいえいきなり第三の実用段階に突入し、ここに音声認識技術一般にブームを呼び起している。それは言うまでもなく TI 社の Speak & Spell である。しかしこれ以前に音声認識技術は特に米国で地道に実用化されていた。それが TTI 社の VIP-100 であり、たとえばテレビ装置の検査マンが読み上げる数値を認識し、それを計算してマージン内かどうかを表示するシステムに利用され

ている。日本では長崎屋が初めて TTI 社の T-600 について NEC の DP-100 を導入し、荷物の仕分けシステムに有効に活用している。これについては本小特集号において日本電気の江袋、磯野両氏による詳しい説明がある。

音声は本質的にオンラインであり、その意味で長所と欠点を持っている。長所は勿論、もっとも自然な人間と機械間の応答手段であるということであるが、一方それは必ず人間の存在を必要とする。(機械音の認識ということも考えられるが、あまり具体的に問題となっていない) たとえば上述の仕分け作業においても、荷物を読み取ることができるロボットがおれば、それですむ話である。すなわちこれは人間の作業効率を上げるための補助手段として利用されるので、どうしても特に価格が問題となる。もっともこのことのためにロボットのごとく社会問題を引き起さない、これは人間と機械の平和共存をはかるものであると言えよう。しかし、仕事によっては二人の仕事を一人にするということができるので、こういう所では経済的に非常にフィットすることになる。実は長崎屋はこの例になっている。このようなことと多分より本質的には後で詳しく述べるその限定された能力のため、特殊用途に限られており、第二の実用段階にあるとはいえない。

こうして見るとパターン認識技術で第二段階にあるのは文字認識技術のみであり、その他は第一段階であり、全体的に言ってまだまだ初期段階であると言えよう。このことは次に述べる実用の内容の所でますますはっきりする。

## 2.2 実用化の質

ここでは、一体いかなる機能が実用化されているのかを見よう。すなわちパターン認識機能を階層化して見て、今そのどこにいるのかを見るということである。音声認識、文字認識、形状認識について表-1 にその階層が示されている。著者の知る限りではこのような階層づけは重要であると言われながら、explicit に述べたものはあまりないようである。ただ音声認識については千葉氏<sup>3)</sup> によるものがあり、ここで示したものはそれを少し補足したものになっている。実際、三つの分野を比べると音声認識が一番明確に階層化されるように見える。特に大事な事は各階層が実用的意味を持っているということである。

これに比べ、文字の場合最高位レベルにある自由書式での日本語文字 (語彙数 2 千程度) というのは実用的観点からすればあまり意味がない。むしろ第 8 レベ

表-1 パターン認識能力の階層化

## a) 音声認識

No.	発生条件	話者	語イ数
4	離散・単語	不特定	100
5	離散・音節	不特定	100
6	連続・単語	不特定	100
7	連続	特定	1,000
8	連続	不特定	1,000/数万

## b) 文字認識

No.	書式	形の種類	文字数
8	記入枠内	常用手書	2,000
9	行記入	(準)自由手書	10/100
10	自由	自由手書	100/2,000

## c) 形状認識

No.	形状の種類	表面状態	背景状態
3	単純図形	非一様中濃度	非一様低濃度
4	不定図形	非一様中濃度	非一様低濃度
5	単純立体	非一様中濃度	非一様低濃度
6	図形	非一様濃度	非一様濃度
7	立体	非一様濃度	非一様濃度
8	任意対象	非一様カラー	非一様カラー

ルの文字枠内、常用手書き、すなわちあるゆるい規則を守ってきれいに書かれた日本語文字という所に意味がある。日本人は欧米人に比べ文字を文字枠内に書くということにほとんど抵抗を感じない。したがって何も困難な自由書式にする必要がないということである。また欧米人の筆記体というのは確かに対象としては存在するが実用的意味は乏しいであろう。しかし、先に述べたごとく音声の場合は、本質的に連続的なのであり、音声発生に区切りを入れるような制限を課することは人間に大変な負担をかけることになる。

一方、形状認識は全体が莫として広く、階層づけが非常にむずかしいが、ともかく最高位にあるものは要するに自然のシーンである。これができれば正に完全なロボットの視覚を与えることになり、一大産業革命を引き起こすことになるが、これはあまりにも困難な課題で、実用的視点からすれば気が遠くなるような話で

ある。

さて三分野の比較はこのぐらいにして、ここで各分野につき検討していこう。まず音声認識では要するに実用化されているのは第一と第二のレベルで、特定話者、単語単位の発声、語イは100程度ということである。この意味で表の背景を黒めにしてある。ここで第一のレベルは実用的観点から見てぎりぎり許される制限で、発声が離散的である、すなわち単語と単語の間を200ms以上発生を休むということで、発声しにくいし、入力速度も落ちる。この点を初めて改良したのが NEC の DP-100 である。これは TTI 社のものに比べ、単語学習が一回で良いという利点を持つ。後者は10回必要である。また実用化のためには厳しい周囲環境に耐えるものでなければならない。これは具体的には騒音であるが、これは一緒に学習することによりかなり雑音が強くなると言われている。この点、ユーザの立場からより具体的な検討が行われようとしている。本小特集号で国鉄の大石、田辺両氏による論文の中でそれが述べられている。

第三レベルの離散的に発声される音節の認識は実用機は一応作られているが、完全なものでなく、また OCR との競争を考えると実用的に問題があると思われる。まだ実用化不完全ということで表の背景を薄めにしてある。

一方特殊環境であるが、電話音声というのは実用的には重要な意味を持っている。しかし音声は高域がカットされるので、機械にとってむずかしい問題となる。この点についても本小特集号において、電電公社の八橋氏はかによる詳しい記述がある。さて電話となると特定話者では話が始まらない。どうしても不特定であることが必要である。この話者の特定と不特定の間には一つの大きな技術的ギャップがある。文字で言えば一定フォントの手書き文字との差に相当する。もっとも不特定とは言ってもそこにまた階層があって、男性、女性、子供と困難になる、これは後者ほどピッチ周波数が高く、正確なフォルマント周波数の推定がむずかしくなることと母音の真の特徴が依然として不明であるということのためであると言われている。

この不特定話者の離散的単語単位発生で比較的性能よいものがパターン大型プロジェクトにおいて、NECにより試作されており、近くその実用機をSR1301として出すということである。一方日立もHR150として実用機を出している。(同上プロジェクトについて成果報告書、文献4)が出ており音声認識以外に

も参考となろう) また通研が精力的な研究開発を行っており、先に述べた八橋氏らによる論文の中で具体的に述べられている。

さて、現在実用化されている単語を丸ごと認識する方法ではどうしても語彙数 200 程度が限界であると言われている。必然的に音素または音節単位の認識方式に移らざるを得ない。そこで初めて連続不特定話者認識への道が開かれることになる。しかし、これは現在あまりにも困難な問題であるということで、タスクをしばり、その範囲で言語、知識を援用してともかく総合力で音声を認識することが考えられ、特に米国で大膽的な研究開発が進められた。これが APRA の SUS (Speech Understanding System) プロジェクトである。しかし、結局わかったことはやはり音響分析と特徴抽出が第一義的に重要であるということであるとされている<sup>9)</sup>。ともかく実用的観点からすれば、かなり絞られたタスクでなければ、ここ当分連続不特定というのは無理であろうと思われる。

次に文字認識について見よう。ここで第二の実用化の意味で実用化されているのは、一定ピッチで文字間に余白ができるように印字された一定フォントまたは印刷されている記入枠内に一つ一ついねいに書かれた手書き文字で、文字種は、英数字、片仮名である。この文字間をあけるということは、音声の場合と同様、セグメンテーションのために非常に重要なユーザへの要請である。さて形の種類が一定フォントであるというのは、実際には ISO で規定された A フォント、B フォントの二種を指す場合が多く、他はオプションとなる。A フォントは機械的で B フォントは通常の英文字の形をしているが、一定文字線幅、一定文字幅 (I などの例外あり) でやはりセグメンテーションへの考慮がなされている。このように人間には、さ細なことで機械にとっては死命を制するほどに重要であるというのは、現状でのパターン認識技術の大きな一つの特性である。このことを理解することはこの技術の実用化を進めていく上で非常に重要であり、鍵である。

また表には書かなかったが文字の品質というのは階層づけの一つの重要な因子で、事実 ISO で印字の品質が定量的に X, Y, Z レンジとして規定されている。品質の良い、X, Y はほとんど完全に、Z レンジもほぼクリアされているが、ユーザは Z レンジ以下を要求してくる。特に最近ではドットプリント文字が良く利用されるようになり、またファクシミリが普及しつつあってユーザの要求は拡大している。実際ファクシミリ

を OCR の直接入力デバイスとして利用するという考えがあるが、簡易入力装置として当然文字品質は落ちる。二値化の閾値を一定に取っているのでもしノイズを防ぐために高めに取れば今度は文字は切れてくる。通常の手書き文字というのはそんなに切れるものではないが、それがかなり著しくなる。現在このような問題を克服しようという努力が続けられており、本小特集号で電通通研の荒川氏らによって、実用的価値の高い、ファクシミリ入力文字の認識が述べられている。

さて手書きについては、自由に書かれた文字を読むというのは数字に限られている。これは技術上の問題もさることながら、英数字を自由に書かれては、人間でももしそれを単独に表示されたらわからない文字対が存在するからである。その典型例が 0 (ゼロ) と O (オー) であり、他の例としては '5' と 'S' などがある。それゆえ、最初米国が、次いで日本が OCR 入力のための手書き文字形の規格を作成した<sup>9)</sup>。実際、ある程度の規則を守って書かれなければ、英数字、片仮名の OCR は実用化できるものではない。この点、幸いにして日本人はよく規律を守るという国民性があり、手先が器用なので、このいわゆる常用手書き文字を読み取る OCR は急速に普及した。労働省の労働市場センターの窓口業務にもこの種の OCR が採用されている。これはいかに規格が実用化に対して重要な役割を果たすかを物語っている。

さて階層 5 は任意のタイプフォントにより印字された文字を、階層 6 は任意の印刷物を読み取るというものである。特に日本では漢字の問題があり、文字数 2 千程度が要求される。英字についてはカナダ郵政省が NEC 製のものを、イギリス郵政省が III 製の、いわゆる Graphics I を利用している。印刷漢字認識技術は PIPS プロジェクトにおいて東芝、日立によって開発が行われ、結局東芝がプロトタイプを作った<sup>9)</sup>。これはまだ実際には使用されていないが、生の特許公報を読み取ることができ充分実用に耐え得るものである。これについては本小特集号において坂井、渡辺両氏により詳しく説明がなされている。

第 7 レベルの最小に規則を守っての手書き英数字、片仮名を読み取る技術は、一層の OCR の普及をはかる上で重要であるが、日本では上に述べたことから、第 2 レベルで読み取りの信頼性、低コスト化をはかるのも、これに劣らず重要なことである。ここで大事なことは、正しく書かれた文字は 100% 確実に読むということ、これは怪しい文字を読み取る以上に大切なこ

とで、えてして怪しい文字を読み取ろうとして、肝心なことがおろそかになることは良くあることである。一方記入枠内という条件の極限として、通常の帳票のごとく、記入枠間余白が零のものは実用的にも、技術的にも重要な問題である。勿論タイミングマークを使えば強制切り出しはできるが、それで問題がすべて解決されるわけではない。

第8レベルはオフラインの日本語入力手段としては理想的なものであり、目下日本のOCR研究者が挑戦する最大のテーマである。研究が始ったばかりで、予想したいが、現在の技術の単なる量的発展上にあるとは思われない。しかし、電総研で行った実験で、単純に位置ずれに強いように改良された相関法で、76%の認識率(きれいに書かれた教育漢字)が得られている。郵便番号の読取りも最初はこの程度であったことを考えると、そんなに困難な技術対象であるとも思えない。ともかく1980年代の格好のテーマであろう。第9レベルのものは日本向けよりもむしろ外国向けのもので、手書きのチェックなどの処理に開発が期待されており、IBMのOCR 3895が利用された例はあるが、コストの割には性能が思わしくなく製造が中止されたということである。これは連続音声認識に相当し、やはり困難な技術課題である。しかし最近NCR(カナダ)がこれに挑戦しているという話を聞いている。

さて形状認識であるが、これは上に述べた二つの分野に比べ、階層化すること、そして実用と非実用の線をごどこに引くか、それ自体大問題である。その理由は対象の多様性にある。ともかく強引にやってみたのが表-1(c)である。ここで最近はやりの動画像などは無視している。

1と2のレベルはいわば人工の場と言うべきもので一定のフォーマットに従って書かれたきれいな図面が1の典型例である。2では対象がそれぞれ分離している必要があり、それが背景状態が一様低濃度であるということに暗に含まれている。もしそうでなければ、音声、文字の所で述べてきたごとく、セグメンテーションという困難な問題を引き起こすことになる。たとえばパーツ(黒っぽい)は白っぽいベルトコンベア上をばらばらになって、一定速度で流れるということが必要である。ともかく、対象と地がくっきりと分れるというのがエッセンスで、これが満足されていけば、いきなり固定閾値で二値化でき、雑音無しの二値図形処理となる。これについては、現在までかなりの技術が蓄積されていて、大ざっぱに言って対象が比較的単純

で2~30種類ならば通常の16ビットマイクロプロセッサで1秒間に数個の形の認識ができる。しかし一応100×100程度の視野で、相関計算のような点ごとの掛算、割算は行わないということを想定している。

しかし現場ではなかなか上述の条件を満足させるのがむずかしい。白いベルトはたちまち黒ずんでしまうであろう。背景も対象も与えられるものであり、そこに改良の余地がない。このような場合にはコントラストが悪くなり、まづいかに対象を切り出すかが最大の問題となる。濃度レベルを豊富(64レベル以上)にすることで、高度の処理を行うということも考えられるが、これはまだ研究室を出ることはできない。一個の対象物を処理するのに数分もかかってしまうからである。これに対してもっとも有効な方法はうまい照明である。ちょっとした工夫により、問題がドラスティックに単純になることがある。たとえば白紙上の白墨を検出するような問題では、まともに上から照明を与え、TVカメラで観測するといった方法では大変である。そこで光ビームを白紙と水平に与えてやれば、光の影が白墨ということになる。しかしこの方法は単純すぎて、たとえばサインペンと白墨を区別できない。もっと一般的な巧みな方法がGMのConsightと言われるパーツの仕分けシステムに採用されている。これは二つの平行ビームを重ねて与えるもので、背景上では一つの直線ビームとなるが物体は入力されるとその上でビームは二つに分離して物体上の端ではビーム光は背景に対し鋭いコントラストを与える。

もう一つの方法は画像を強調することである。特にこれは不定図形として考えられる、傷、欠陥等の検出に有効である。日立のプリント配線の欠陥検出はこの良い例になっている。GMのSight Iでは、微分を行うことにより、ICチップと熱源のコントラストを強調している。ともかく与えられた条件を最大限に利用し、目標を絞り、効率良い方法を見出すことがエンジニアに課せられている。

このように、実用になるかならないのかは、対象とその置かれた環境に強く依存し、個別的である。第6のレベルになると、非一様濃度であるため、セグメンテーションの問題と正面から取り組まざるを得ず、現状では実用化には遠い話となる。これについては一番大変な切り出しを人間がやり、切り出された形についての物理量(面積、重心等)の測定は計算機が受け持つというマンマシンのシステムが考えられる。これは完全な省力化にはならないが現在の技術水準に合った

現実的アプローチであると、現場の技術者は見ている。

以上三分野について、実用の質を見てきたわけであるが、共通して言えることは、対象が分離していること、対象と背景のコントラストが良いこと、対象の種類が百程度（印刷漢字を除く）であること、対象の変動はできる限り少ないこと、等である。そしてこれらの具体的対象は人間の接するパターンの世界から見れば、ほんの一部にすぎない。そしてそれは時として、不自然に作り出された対象でさえある。すなわち人間と機械の妥協が必要であり、これが重要な課題となっているということである。

### 3. アルゴリズム

以下アルゴリズムとハードウェア、そしてアルゴリズムの代表的なものとして、マッチングアルゴリズムについて概観する。

#### 3.1 アルゴリズムとハードウェア

認識アルゴリズムの数は多い、しかしその原理は共通していて、しかもきわめて単純である。すなわち‘マッチング’である。特に実用的観点からすればこのように言い切っても良いであろう。ただこの原理が具体的に適用される対象によって、またそれを実現するハードウェアによって形を変えるのである。

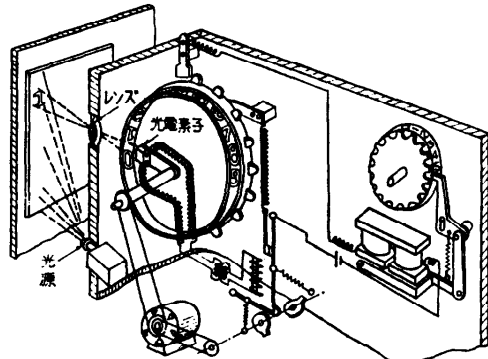
もちろん、この原理の具体化は時代と共に進歩している。特に興味あるのはハードウェアの進歩と歩調を合せているということである。いわばここに実用指向アルゴリズムの宿命を見ることができよう。

図-1はこのハードウェアの進歩をOCRを例に挙げて並べたものである。(a)は機械光学的方法であり、エレクトロニクス時代の前夜、1927年に世界で最初のOCRとしてTaushekにより考案され、特許となったものである。ここに、パターン平面でのいわゆるダイレクトマッチング、またはテンプレートマッチングの原理を明確に見ることができる。当時の技術的状况からすれば、これが多分唯一の解であったらうと思われる。

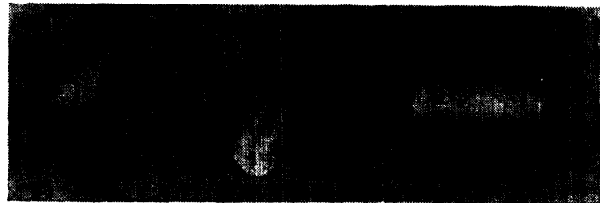
ついでエレクトロニクス時代の幕が開き、真空管時代が開花した。(b)はその雄であったRCAの技術者が考案したものである。(a)と原理は同じである。ただネガの考えがはいっていて、たとえばEの入力がFのマスクに一致することを防いでいる。これも真空管時代の当然の帰結であると言えよう。しかし高度な技術を駆使するあまり、一つの華麗なる記念碑に終った。

(c)は超高性能電子計算機の大型プロジェクトで電総研と東芝の協同開発になるOCR、ASPET 70の複合類似度法によるマッチング回路図である。技術者はこの時代に来て初めてアルゴリズムの自由度を手に入れることになる。アナログ、論理の電気回路を駆使することにより、文字通りのテンプレートマッチングから、なんらかの特徴を抽出してからマッチングができるようになり、種々の特徴抽出のアルゴリズムが考案された。しかしたとえば典型的に使用される演算である係数を掛けて加算する演算に、この時代ではアナログ加算器が好んで使用されたが、その精度（特に音声において）の悪さと、配線の煩雑さのため見捨てられ、かくしてLSIの時代を迎えることになる。

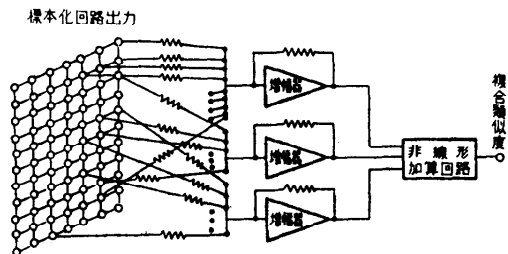
(d)はこのLSIの一例としてパターン情報処理シ



(a) Taushekの世界最初のOCR原理図



(b) RCAによるマルチフォントOCRのマッチング機構



(c) ASPET 70の複合類似度法によるマッチング回路

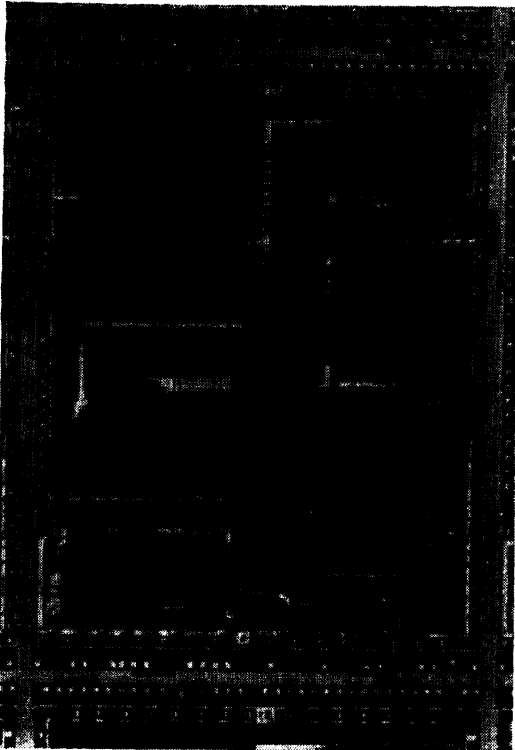
(d) PIPS プロジェクトで開発された  $\mu$ P, PULSE

図-1 OCR ハードウェアの進歩

システム大型プロジェクトにおいて開発された高性能マイクロプロセッサである。実際電総研において、これを用いいわゆる ONE- $\mu$ P-OCR を作った。ともかく抵抗は ROM に、加算は  $\mu$ P の CPU に、そして配線プログラムによってかえられたのである。ここに至って技術者はアルゴリズムに大きな自由度を得、気がねして使ってきた掛算とか割算、それに対数変換といった高度な非線形演算を採用できるようになった。メモリが高速化しかつ安価となり、標準マスクの数も今や 100K ワードとなっても別に驚かなくなった。だいたいメモリはある意味では演算を代行するもの（変換テーブル）であり、またマッチング法を実現するにはもっとも直接的な媒体である。もっとも、現場の技術者はそんなに自由であるわけではない。LSI 回路をいかにけちるかについては Speak & Spell のその音声合成回路の設計に如実に見ることができる。この意味で技術は常にその時点における cost-performance の極限を目指すものなのである。

結局、現在の実用化をもたらしたものはいわゆる丸

ごとのマッチングアルゴリズムとそれを支える LSI 技術であるということになる。

### 3.2 マッチングのアルゴリズム

前節で用いた丸ごとのマッチングという言葉をここでもう少し厳密に考察する。ここで抽出された特徴は時間または空間上の座標を持っていて、それらの関係は暗に時間または空間上での位置関係として与えられる。そしてまったく同じ表現が標準パターンについてもなされ、マッチングはひとまとまりの入力パターンの特徴表現と同じくひとまとまりの標準パターンの特徴表現との間で、その両者に共通する空間上でなされる。なお、二次元パターンの場合でも、必ずしも二次元空間上でマッチングとは限らない。実際、対象の輪郭を追って特徴を出せば、一次元となる。

この意味では時間という一次元上の現象である音声の場合は順序が一意的なので、この丸ごとのマッチングが比較的容易である。また二次元でも、文字はほとんどの場合、正常位置にあることを仮定して認識が行われる。要するに回転があまりないということである。それゆえ、文字の大きさを正規化すれば、特徴の位置関係がかなり順序づいたものとなり、丸ごとのマッチングが有効となる。人間はこの点あまり気にしないが、機械には非常に重要な条件となっている。これも実用的に有効な制限の例である。

ところが一般に物体の場合には、そうはいかない。上下左右という絶対的位置というものが意味を失っていく。信頼できるのは相対的位置関係のみである。ここで座標という自然の順序を頼りにした丸ごとのマッチングは力を失い、正統的な二次元上のマッチング技術が本質的に要求されることとなる。

さて丸ごとのマッチングでも、実際の所はそんなに簡単ではない。位置の正規化が重要になってくる。位置ずれによる相関のロスを防ぐため、ほかしの操作などが工夫されている。特に音声では時間軸での正規化を必要とし、このため入力と標準との対応が、順序保持の条件の下でより柔軟に行われる。しかしこの場合、対応関係の組合せが一気に増大するので、その発散を防ぐため DP (ダイナミックプログラミング) が採用され、これが成功している。同様な考え方を二次元空間上に拡張することは原理的に可能で、実際相当に効果のあることが手書き数字について立証されたが、同じ考えを英数字、片仮名に拡張すると、大変な手数、計算を必要とし、正に実用的な意味でこの方式

は実用機には採用されなかった。

この丸ごとのマッチングでは、まず丸ごとにする必要、すなわちセグメンテーションが非常に重要な要件となる。しかし、特に音声の場合、対象は本質的に連続であり、この点、セグメンテーションに若干の可能性を与えて組合せの増大をまた DP で吸収するという手段が取られている。これが2段 DP 方式と言われるもので、連続単語認識で実用化されている。これに反し、いわゆる流し読み方式で入力部分が現れるごとに、いわば連続的にマッチングを取る方式が考えられ、特に音声では時間という方向の流れが明確なので、これが有効である。この連続 DP 方式は LSI 化にも適しており、実用化に有望な方式と言われている<sup>6)</sup>。

一方、丸ごとのマッチングを離れて、入力、標準(モデル)間の特徴の対応をある程度自由に考えなければ、任意の位置にある物体認識は一般的にはできなくなるということは前に述べた。これは抽象的にグラフ上でのマッチングを取ることになる。しかし対応関係の組合せが非常に増大するので、これをしぼる種々のアイデアが試みられている。二次元 DP もその一つであり、問題解決の分野で開発された state search algorithm, グラフ理論での maximum clique, そして最近よく研究されているのが, relaxation method である。しかし、これはパラメータの選定と計算時間に問題があると言われている。とはいえ、時間の問題については方式が本質的に並列であり、VLSI の到来と共に実用的価値を持つかも知れない。実際この方式の装置化の研究もなされている。

いずれにせよ上述の高度のマッチング方式はまだ研究室の域を出ていない。これらが部品技術の進歩のどの段階で、実用化されるのか、またはこの他により良い実用的マッチングアルゴリズムが存在するのか、非常に興味ある問題である。

#### 4. あとがき

これは解説ということになっているが、私自身はそ

のつもりで書いたものではない。パターン認識の実用化という問題について一つの見方を与えることを意図したものである。その意味で、一つの論文と見るべきものであり、またこれ自身が問題提起となっている。実際、ここに述べたことそれぞれが非常に大きな問題でとてもこの紙面でカバーできるものではない。したがってこれは未完成品である。そのつもりで読んでいただきたいと思う。

またこのような企画が PIPS 大型プロジェクトの終りになされたということは非常に意義深い。関係者の一人として、心から喜びたいと思う。

#### 参 考 文 献

注：以上の考察は音声、文字、形状というパターン認識技術のほとんど全分野にまたがっており、まともに文献をリストすればそれだけで膨大なものとなるので、最小限に限った。ただ一般的なものとして以下の本を挙げておく。

\* 坂井、長尾：文字・図形の認識機械、共立出版(1967)。

\* 中田和男編：パターン認識とその応用、コロナ社(1978)。

\* 中田和男：音声、コロナ社(1977)。

\* 白井良明：コンピュータビジョン、昭晃堂(1980)。

\* Rosenfeld & Kak: Digital Picture Processing Academic Press (1976)。

訳本：長尾 真監訳：デジタル画像処理、近代科学社。

1) Sugarman, R.: The blue-collar robot, IEEE Spectrum Vol. 17, No. 9, pp. 53-57 (Sep. 1980).

2) Computer, Vol. 13, No. 5 (May 1980).

3) 千葉成美：単語音声の認識、情報処理、Vol. 19, No. 7, pp. 667-674 (1978).

4) 大型プロジェクトパターン情報処理システム、研究開発成果発表論文集、電総研、PIPS-R-No. 25 (昭和55年12月)。

5) 中田和男：音声の合成と認識、総合電子出版社。

6) JISC 6254, JISC 6255 日本規格協会。

(昭和56年2月2日受付)