

コンピュータビジョンの将来技術での位置づけ

坂井 利之

龍谷大学 理工学部

コンピュータビジョンについて、その将来技術の中で果たすべき役割と位置づけについて述べている。動物、人間に備わっている各種感覚器官の中での視覚センサーと、その外在化としてのテレビジョン、さらに一般的に各種信号波の可視化・カラー化、さらには人工環境空間・バーチャルリアリティによるビジュアルメディアの拡大・変革を考える。

次にマイクロマシンに代表されるように、ロボットをはじめ、各種の機器・システムの細胞化・インテリジェント化の中で、高速化、高度知能化が必須の条件になることを示す。

コンピュータビジョンが、このような技術趨勢の中で、もつべき機能について考察し、さらには情報処理として状景表現単語、感性的表現のマシン記述化が、今後の大きい研究・開発動向になるであろうという意見と例題を示している。

Situation of computer vision in future technologies

Tosiyuki Sakai

Faculty of Science and Technology

Ryukoku University

1-5, Yokotani, Ohe-cho Seta, Otsu-shi, Shiga-ken, 520-21, JAPAN

In this paper, situation and target of computer vision is described from view point of future technologies.

It is compared visual sensor among various human and animal sensors with artificially visualized images, color representation technique of various signals.

Moreover, artificial environment space is considered as extended, integrated space of real world and virtual reality spaces.

It is also mentioned necessities of micro-machine, real time and highly intelligent processing.

Machine understandable description on various emotional scene and sight of human face, human behavior may be very difficult, but would be essential in multi media network societies.

1 主体とセンサー

1.1 動物的感觉

ここで言う主体とは、生物、動物であり、また、センサーとしては動物的感觉、すなわち動物の生命の保全、本能の満足のために備わっているものをさす。

しかし、人間は、動物的な感觉と対称的なものをも持っていて、それは思想や意図と言うものになる。これは単なる生命の保全、本能の満足と言うものと異なって高度なものと考えられている。

1.2 視覚センサー

次に、視覚センサーについて考える。視覚センサーは、指向性を持っており、ある方向を見ようとする主体の意志の制御下において、はじめて機能を発揮する。それに対して、音声は無指向性で、意図に関係なく情報が入ってくる。聴覚系は、たとえ、眠っている場合にも大脳の最終段階直前まで動作している。視覚センサーの非常に大きな特徴は、遠隔的な状景でも非常に高速な光というものを用いて、遅延時間が少なく粒子性あるいは波動性（ホログラム）として把握することが出来る。これは、音波や超音波等のように、非常に伝播速度の遅いものと比較すると、遠隔のものを捉えようとする際に、音波では遅延が起こる点と非常に違った特徴になる。

2 可視化・カラー化・環境化技術

2.1 各種波長光による人工擬似画像

まず、各種波長光による人工的な擬似画像について考えてみるが、以前は自然の光、すなわち可視光と呼ばれるものによって人間や動物が視覚系で外界を把握することを行っていた。最近ではあらゆる波長の反射波（到来波）、例えば、超音波の材料、自然界でのエコー、あるいは人体の臓器からの超音波のエコーの強弱というような種類のことを、全部擬似的に像の濃淡に変換して画像としてカラー化することが出来る。これは、電波レーダから始まっているように、人間感觉の残像現象を利用して、環境をイメージ化して測るということであったが、超音波のときは、ソナーと呼ばれている。白黒の画像がその後、濃度値によりカラー化に進んできた一方、現在は、 n 枚の中から選ばれた3枚の同一サイズ、同一パターンのいわゆる疑似カラー画像がリモートセンシングで多用されている。

2.2 生物的感觉・センサーより広範囲のコンピュータビジョン

次に、生物的感觉・センサーよりは広範囲のコンピュータビジョンについて述べる。一般の自然環境における生物的感觉、センサーに比べて、コンピュータビジョンというものは各種の波長、す

なわち音波から γ 線まで、様々な種類のものを並列に利用することが出来る。刺激波をこちらから発射して、その反射波または透過波を捉えるという方法に対しては、波長の極めて短いものから長いものまであらゆる種類の粒子波、光、電磁波だけに限らず超音波の範囲のものまでを全部可視化、カラー化出来るということが最近の大きな特徴である。

こちらから刺激波を送らず、生物、自然界から放射される各種波長の受信、可視化も当然存在する。

2.3 人工環境空間の形成と学習・行動（VR）

以上を総合して、人工環境空間の形成と学習・行動（VR）について述べる。人工環境空間というのは、様々な受信波、あるいは発射した刺激波とそのエコーというものによって、我々は自然と人工的な空間をそれぞれあるいは融合して形成し、その状況を観察することが出来る。これは動物が備えている極めて限られたものと比べると、非常に広範囲で柔軟な空間になってくる。そして、これによって環境空間というものを探索する能力は、通常の動物よりも著しく優れたものになると言える。その極限というか、現在の技術での極限は、いわゆるバーチャル・リアリティー（VR）いうことになる。

今まで人間が自然界で観測することが出来る範囲外のもの、コンピュータで人工的に作るもの（現実には存在しないが機械の中で出来るもの）それと自然環境やそのイメージとをドッキングさせてその中で人が、ロボットが学習していく、あるいは行動を起こす。

相互に密接な関係をもつ主体を含んで、それらを内包する統合的な空間を形成することが出来ることになる。それは人間には危険な実空間でも、娯楽空間、ゲーム空間にも、あるいは遠隔空間とのドッキングにも利用可能なものとなろう。

3 ロボットのインテリジェント化

3.1 コンピュータビジョンの各種センサー機能中での比重

まずはじめに、コンピュータビジョンの各種センサー機能中での比重について考える。現在各種波長のセンサーが存在する。触覚、聴覚、あるいは視覚、嗅覚、味覚と様々なものがあるが、電気変換の実現、普及しているのが前の3つである。

これらの中でコンピュータビジョンというものがどういう比重を占めているかということ、遠隔のものまでセンスすることが出来、またそれをキャッチ出来るのはオーディオとビジュアルである。触覚は直接接触らなくてはならないし、音波は無指向性であり、視覚はビームのように指向性でそちらの方向を向かなければ分からない。勿論、魚眼レンズのように広範囲のものを把えるものもありうる。

味覚や嗅覚の器官は、体内に存在するものであるもので、空間はさらに至近距離というか、限られたものになってくる。

コンピュータビジョンというもの、あるいは視覚という種類に関しては、遠隔のものを遅延時間を少なくして空間のある方向に限りセンスするという点では、もっとも優れているというか唯一と考えられる。

3.2 観察→判断→行動のサイクルのリアル・タイム性

次に観察→判断→行動のサイクルのリアル・タイム性について述べる。例えばロボットの場合についていうと、周囲の環境と対象物を観測し、それに対して適切な判断をする。そして、その判断の結果に従って行動するというのを、ある許された時間範囲でリアル・タイムに処理するということが必要になってくる。勿論、行動を起こせば世界も変わるということになるのでサイクルという言葉を使っている。

今現在、通産省で行なっているプロジェクトでいえば、リアルワールドコンピューティングというようなことになろう。

ロボットはパターンや世界を観察する。そして目的に沿った問題解決をし、行動を起こすということを、少ない場合ではms程度、長くても秒単位で処理しなくてはいけないということになる。従って上記サイクルの一連の行動が超高速に行なわれる必要性が生まれてきている。それが例えば原子炉の場合やナビゲーション、防災など様々な場面においてリアル・タイム性を必須とするタスクによって、非常に限られたコンピュータビジョンになってくるということである。

3.3 マイクロ・マシンの不可欠さ…知能感覚器の分散配置から細胞化へ

マイクロ・マシンの不可欠さ、すなわち現在VLSIというものが生まれてきているということは、単に小さい所に全てのメモリや処理装置の機能を入れようとする努力、あるいは軽薄短小ということだけでなく、同時にメモリもプロセッシングの機能も全部小さいところに入るということで、知能のあるマイクロ・マシンの出現ということになる。マイクロ・マシンというと、当然電源やモータ等の動く装置なども全て小さくなっていく必要があるが、これらは知能の付いた感覚器というふうに言えるであろう。それがマイクロ化されるということは、あらゆる所に置くことが出来るということで、装置や伝送路のどこにでも分散配置し、その細胞自体が一つの頭脳と行動力を持った自律的、動的なマシン（マイクロ・マシン）といえるであろう。

3.4 巨大人工動物の誕生

このように考えてくると、今までのあらゆる動物と比較して巨大な人工動物が、現在では至る所に存在することになる。それは例えば、神経の伝送系や運動の制御系、判断中枢というようにはっきりしたものに集中化されるのではなく、頭脳や行動系、あるいは感覚器官というものが融合し-

体となって、至る所に存在する巨大な人工動物というものが生まれているというのが、実はネットワーク社会ではないかと思うのである。

4 コンピュータビジョンのもつべき機能

4.1 各種波長波のビーム化、エコー波の変換・可視化

各種波長波のビーム化、エコー波の変換・可視化ということが、一つの必須の機能として求められる。すなわち、各種波長の刺激を、出来るだけある方向にビームとして出すということ、ビームとして出せないときは、例えばCTのように、ビームになっていないものをあたかもビームにしたかのようにする技術も含まれるが、その透過波やエコー波というものをそれぞれ目的に応じて変換し、可視化して、それを2節で述べたように利用することが、最小限備えていなければならないことになる。

4.2 疑似画像の多重化、差画像の検出・符号化

疑似画像の多重化、差画像の検出・符号化の説明をする。カラー画像ということになると、3枚の同じパターンの各色による重畳が必要になる。マルチスペクトルバンドのエコーに関しては、例えばマルチバンドのビームとして、チャンネル数8あるいは5といった同一サイズ、同一対象物のエコーパターンの中から3つを選び出して、カラー化が出来るということになる。

すなわち、3色(3次元)よりもはるかに大きい多次元の空間において、我々は対象物の多次元の画像というものを観察することが出来るようになってきている。動画においては、前のフレーム(フィールド)と後のとを引いてみて、その差を検出することが移動体の検出に非常に重要であり、また動画画像の圧縮の点でも極めて重要である。すなわち画像技術は、単に情報処理、パターン認識というだけでは無しに、通信や蓄積技術とも関連して、益々重要な地位を占めてきていると思われる。

4.3 画像処理の高速化・超高速化…ナビゲーション、ロボットシステム

次は、画像処理の高速化・超高速化ということである。これは前節のロボットの所でも言ったように、リアルワールドコンピューティングで象徴されるようにパターン認識あるいは画像処理が超高速に行なわれる必要があるということである。

例えばナビゲーションという場合には、今現実には主体である物体が移動している時間内に、その移動に必要な情報を、周囲から観測し、摂取し、そしてそれに対する判断が必要になってくるということで、知能ロボットシステムと同様に高速化は当然不可欠な機能となる。

それから、話しは変わるが、歯科医師が、患者に対して様々な処置をしている場面を考えると、患者・付添と歯科医師・補助員とが、現時点が治療のどの作業状態にあるのかどうかによって、治療

空間の状態が幾つかに分類される。その状態をマシンビジョンによって判断し、歯科医師が患者に対して事細かな指示をしなくても、勝手にマシンが刻々と状況を判断してくれる、そして患者に対してマシンのほうが音声で指示をするということが可能になると思っている。その時の音声は日本語に限らず、また言葉の話しぶりも選ぶことができる。

4.4 画像からのファジイ、感性情報の抽出、類別

画像からのファジイ、感性情報の抽出、類別ということは、これからは特に重要な研究テーマであると考えている。我々にとって一番大事なことは、今のコンピュータビジョンでは判断出来ない画像ではあるが、人なら楽に判っている画像に対して処理をする、ことである。例えば顔画像を見て、男なのか女なのか、老人なのか幼児なのか、あるいは太った人であるか痩せた人なのか、また顔の表情はどうか、困ったような顔であるか嬉しい顔であるかという種類の非常に感性的なもの、感情の入った種類のもを抽出したり類別するということが非常に重要になると私は思っている。

ここでマルチメディアというようにことを例題にしよう。光ファイバーや人工衛星によって、非常にたくさんの画像チャンネルがとれるようになってきているが、自然言語というものが世界には3千数百もあり、それがネックになって世界中の情報流通が妨げられる。ゲームのように、ある限られた約束の下にあるものは、言語が通じなくても解釈できるが、それにしても自然言語は多少は入り、それが分からなければ、この壁は依然として強いのである。

これをなくそうとする際に、人がいちいち文字を入力したり吹替えをするということは、24時間態勢で、チャンネルが多いと到底出来ないわけである。画面上でマシンが顔を見た直後に、男の人の場合は、男の人の声で翻訳した文章を使う。またそのときに大人であるか子供であるかという事を、画像の中からマシンが読みとって、それにふさわしい対応した声や顔で、それぞれ処置をリアルタイムにすることが理想である。マルチメディアを叫ぶのなら、これからの非常に重要な感性情報処理の原点は、ここから始まっていくのではないかと思うのである。

5 画像理解の高度化・感性化

5.1 状景表現単語のマシン記述化

まず最初は、状景表現単語のマシン記述化ということである。我々の自然言語には、各種の状景を表現するような単語が色々ある。画像処理でよく用いられているテクスチャというものは、状景表現単語の画像内での定義・分類であり、何かほぼ周期的であるとか、ほぼ対象に近い等、非常によく似ているものの配列になっている。

例えば、ひび割れを見て、これはどういう形のひび割れであるのかというようなことが、出来てきたとしたら、それが材料であろうが壁であろうが、大変な技術進歩をもたらすことになる。動

きのあるものとしては、“挙動不審”とか、“熱心にショーウィンドーの品物を見る”というのもありうる。

こういう種類の各種表現単語を探してみる。そして、その単語に対する機械の中での記述を研究することが必要になるのではないかと考える。

5.2 ゲシュタルト心理学特性のマシン記述化

もう一つ、同じような形でいうと、心理学、特にゲシュタルト心理学などで指摘しているように、これは滑らかなものであるとか、これはお互いによき形であるとか、同じような方向とか回転であるとかその輪郭とか、心理学上でいわれている近接性、類同性などの特性というか画像の形とか法則という主観的に近いものを、機械の中で記述するという研究が求められると思うのである。

5.3 産業界・医学界等の同属・多重画像の自動モデリング

次に、産業界・医学界等の同属・多重画像の自動モデリングということが必要であると思う。現在、製造産業界では、出来上がった製品を抜き取り検査しているのが多いが、これでは遅い訳で、例えば布地、鉄、紙等を作っている行程の中で、製品状況を見れるようにしたいのである。医学界でも、医療に関しても、最近では画像が至る所に使われており、こういう画像パターンにおいて、似ているか似ていないか、同じ種類に属するか属さないかというようなこと、あるいは各種の画像というものを自動的にモデリングしていく研究というもの非常に重要になってくる。例えば、心電図でいうとP波やQRS波、地震についても脳波についても同様で、分野ごとに様々な波形が定義されている。モデリングということ、画像だけに限らず機械から発する波形、生物や自然界が発する波形についても、もっと定義され、それについての研究を進められていく必要があると考えるのである。

5.4 正常、異常（非常、警戒）の判定の重要性

次は、正常であるか、異常であるかの判定の重要性について述べる。異常というのは非常事態にあるのか、警戒状態にあるのかということで、要するに信号というのは時系列、つまり時間の関数であるが、詳細な状況のモニタリング画像をテレビカメラや各種のセンサーでとり、光ファイバーで幾らでも遠くへ送ることが出来るけれども、最後には、人間でなくマシンが異常さを判定する必要がある。今現在の状況が、これはなんの異常も無いというノーマルな状態である時には、情報を送ったり、記憶する必要は全く無い筈である。

問題はノーマルな状態、あるいは我々がなんの関心も払わないというものに対しては素通りさせておいて、非常におかしい、あるいは重要な問題を起こしそうだというような異常とか警戒すべき状態というものがあった時に、その状況画像を撮る必要がある。

そういう意味から言うと非常に長い時間の中で、非常に少ない異常状態、警戒状態を判定するということが必要になるわけで、換言すれば、これは人間だけでは不可能なことで、機械にさせざるを得ない問題である。

何ら変化の無いもの、あるいはノーマルな状態では何もしないでパスさせるという判断をするマシン、これは私が前に情報フィルターと呼んでいるものの一種になるが、そのようなものが不可欠ということである。

5.5 マシンモデル適用内と適用外れの告知能力

マシンモデル適用内と適用外れの告知能力とはどういうことかと言うと、我々はあるマシンを使っている時に、そのマシンのプログラムあるいはそれがどういう条件の下には“成立する”、どういう条件の下には“成立しない”か、という根本原理が存在するが、それを常にチェックする機能である。

一度使い始めると、その範囲を外れている場合でも、無造作に適用するということが往々にして起こるのである。これと似たことは、有効数字でいつも言われることであるが、モデルをマシンの中に入れているときに、今、それを現に適用しているものは、モデルを作成する時に考えていた条件・範囲に入っているからOKというふうになると、そろそろこれはおかしくなって適用の範囲を外れているというような事を知らせないと、最終的に何をしているか分からないことになるという注意である。

画像理解や高度な問題を扱う時には、根本原理として条件に入っている前提とか、モデルを作る時の仮定を、いつも認識しチェックして、これを告知するような種類のものを導入することが、コンピュータビジョンの分野でも当然必要になってくるのである。

6 あとがき

コンピュータビジョンそのものの狭い範囲だけの将来技術での位置づけでなく、広く情報処理一般、そして、情報処理と電気通信、電子情報記憶メディアとの融合の立場から記述してみた。会合の席では、この研究会によりふさわしい例題とか話しに焦点をしばって述べる積もりである。