# 最近のコンピュータービジョンの動向

村　瀬　　洋

コンピュータービジョンとは，人間が視覚の情報から外界の状況を認識・理解するような情報処理機能を，機械により実現することを目的とした研究分野である．以下CVと呼ぶことにする．この用語は，狭義には3次元物体や外界の3次元構造を理解する場合に限って使用されることもあるが，一般的には画像認識や画像理解等に幅広く用いられている．その歴史は古く，1960年代には文字認識等のパターン認識やディジタル画像処理（2値化，細線化），1970年代には3次元情報処理（ステレオ），1980年代には多次元情報処理（shape-from-X，動画像処理，レンジセンサー）などが大きなトピックとして取り上げられてきた．特に1980年頃に提案されたMarr[1]の視覚の計算理論的なパラダイムは，CVの研究方法に大きな影響を与えた．その影響を受け，CVの問題は様々な部分問題に分割され，各々の部分問題に閉じた形で研究される傾向が強まった．しかし，このアプローチに従った研究は幾何学，光学，最適化問題などの手法を駆使し，学問的には大きな成果を上げたものの，実際の環境の中でロバストに利用できる認識・理解のシステムがその延長線上で実現されるためには，まだ時間がかかりそうな状況である．近年着目されているアクティブビジョンやアピアランスの基づく認識手法は，これとは別のアプローチであり，部分問題に閉じることなく全体の目的を考え，それを直接的に実現するシステムを構成していこうとする手法である．

CVの研究については長い歴史があるだけに，これまでも学会誌などにいろいろな観点から解説が掲載されてきた[2-5]．さらに，多数の優れた教科書的な本[6-12]が出版されており，個々の手法についてはそれらを参照することが望ましい．図1に画像理解の分類を示す．本解説では，頁の制約などもあるため，最近の傾向および著者らが進めている固有空間法による画像認識手法を紹介するにとどめる．

## 1. CVをとりまく背景

### 1.1 計算機の進歩

画像を処理するには一般にメモリーや処理能力などの計算機の資源を非常に多く使用するために，これまでのCVの歴史は計算機性能の進歩に大きく依存してきた．しかし近年，計算機の性能向上は著しく，100Mバイト以上の実メモリー容量，100 MIPS以上の計算能力の小型計算機も身近なものとなっている．近い将来には，この値はそれぞれさらに1桁以上向上していることであろう．今後のCVは，従来のように計算機性能の制約に縛られることのない，より自由な発想の手法が生まれてくるはずである．逆に，研究者はこれらの計算機資源を有効に利用していくべきである．

### 1.2 マルチメディア環境の普及

計算機の小型化，低廉化，高性能化とともに，その周辺機器，表示機器等も身近なものとなってきた．例えば，最近のマルチメディアパソコンではビットマップディスプレイにより手軽に動画像を表示する機能はもちろん，画像や動画像の取込み機能が標準で組み込まれている．これらのものを個人が一般の電気店で購入できるレベルまでマルチメディア環境は一般化している．ビデオカメラやビデオレコーダーも一般家庭に多く普及してきた．つまり，高価なフレームバッファーが購入できる大

Trends in computer vision（1995年12月27日受理）
Hiroshi MURASE NTT基礎研究所（〒243-01 厚木市森の里若宮 3-1）

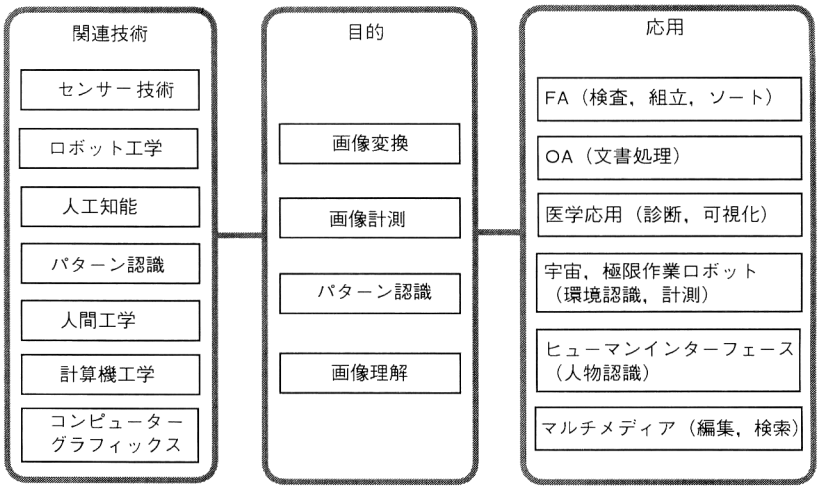



図1　コンピュータービジョンの分類.

きな研究機関でしか画像の研究はできないという時代は完全に終わった．今後これらの安価なマルチメディア環境を生かしたCVの研究が大きく発展するはずである．また産業界においても，これまで画像機器は高価であるという理由から利用を敬遠していた分野においても，システムの低廉化により導入の可能性が出てきたわけである．

## 2. 応用からみた最近動向

### 2.1 マルチメディアにおけるCV

最近着目されはじめたのが，マルチメディア環境での画像認識技術である．前節で述べたような個人のレベルで収集された映像に加えて，インターネットなどを通してディジタル化された映像情報が身近にあふれている．そこで，画像通信・符号化，映像の編集・操作の要求が自然に生まれてくるわけであるが，この分野においてもCVの技術は有効に利用できる．

(1) 映像情報の構造化[13]

映像の構造化は，映像を操作する上での重要な要素技術である．構造化には，(i)空間的構造化（画像の張合せ，物体の位置関係の抽出など），(ii) 時間的構造化（シーンの切れ目の検出など），(iii) 意味的な構造化（画像に含まれる意図の抽出など）がある．これらを自動的に行うにはCVの技術に頼らざるを得ない．

(2) 画像検索[13,14]

映像データベースから特定の内容を含む画像を検索する技術も着目されている．これはcontent based retrievalと呼ばれている．色，形，テクスチャー，さらには音声などの他のメディアも併用した様々な特徴を利用した検索手法が検討されている．

(3) 画像通信[15]

最近の動向としては，インターネット上の画像等のデータベースアクセス（mosaic，netscapeなどのこと）に適したインタラクティブ符号化などが着目されている．さらには，3次元統合通信という枠組みでは，光線の情報を $f(x, y, z, \theta, \phi;\ t)$ ($x, y, z$ は光線が観測面と交差する座標，$\theta, \phi$ は光線が観測面と交わる角度，$t$ は時間，$f$ は光強度を表す関数）で表現する．これはCVの表現法ときわめて類似している．このように今後ますますCVと画像通信の境界は曖昧になってゆくであろう．

### 2.2 医用画像処理

もう一方で，医用画像処理[5]はニーズに裏打ちされながら画像認識技術と結びつき，着実にその研究領域を拡大してきた．X線CT装置（最初は1972年英国EMI社より発表）の発明以来，医用機器と画像処理との関係は深い．自動診断分野では，スクリーニングや特徴量の数値化等を用いた研究が進んでいる．それ以上に医師にとって特に大きな助けになっているのは，ビジュアライゼーション技術である．その中でもボクセル表現等は，メモリーを大量に使用するものの画像の操作が簡単なため，広く利用されるようになった．その理由は，前節でも述べたように，大規模なメモリーを容易に利用できる計算機環境の変化によるところが大きい．

## 3. 手法からみた最近動向

CVの手法は，計測，検査，3次元復元，認識，理解，符号化など非常に多岐にわたるために，最近の動向を一般的に述べることは困難である．ここでは，情報処理学会のコンピュータービジョン（CV）研究会で最近取り上げられたキーテクノロジーを紹介する．当研究会では，インパクトの大きい話題性のある手法について将来動向を議論している．現在までに取り上げられた手法は

(1) Hough変換[16]
(2) 不変特徴量[17]
(3) アクティブビジョン[18]
(4) 正則化[19]
(5) 多層ニューラルネットワーク[20]
(6) 固有空間法[21]

であり，今後もさらにいくつかのテーマが取り上げられてゆく予定である．ここに登場する各テーマが最近の手法の動向を示しているものと考えられる．ここでは，(3)，(4)，(6)およびphysics-based visionについて簡単に述べることにする．

### 3.1 アクティブビジョン

アクティブビジョン[18,22]はCVにおいて近年着目されている一分野である．アクティブビジョンは，例えば計測等，そのシステムの目的を実現するために都合がよくなるようにカメラの位置などを計算機で制御しようとする手法で，ロボットビジョンの研究と親和性が高い．また認識・知覚のモジュールと行動のモジュールの統合技術とも考えられる．さらに発展して考えれば，これは視覚情報処理の研究というよりは，目的を達成するための系全体の研究ともいえる．

カメラを計算機で制御するという発想は，1960年代後半のSRI（Stanford Research Institute）等で行われたハンドアイシステム[22]などにみられるが，最近Bajcsy[23]，Aloimonos[24]，Ballard[25]がその重要性を改めて説いたことにより再び着目されている．その理由は，1枚の画像だけでは困難なCVの問題が，カメラを動かすことにより簡単になる場合が多いからである．Aloimonosが初期に定義したアクティブビジョンは，センサーの幾何的パラメーターのいくつかを運動パラメーターに置き換えることにより，従来解きにくかったビジョンの問題を解こうとするものであった．彼はその後この考え方を発展させ，定性的な情報を獲得するモジュールの集合としてシステムを構築するpurposive vision[24]へと考え方を発展させている．Ballardの提案したanimate vision[25]では，文脈に応じて視線を制御する注視制御の重要性を説いている．アクティブビジョンの研究はその後さまざまなバリエーションが提案され，最近ではモデルや環境を獲得するための注視制御の手法などが研究されている．しかし，これらの注視対象の選択や注視点の制御は，一般的にタスクに依存する困難な問題である．そこで人間のノウハウをロボットに教え込もうとする研究もある．また人間の視覚システムを参考にしたようなセンサーシステムの研究もなされている．

### 3.2 正 則 化

CVでは，入力データからのみでは解が一意に決定できない不良設定問題が多く存在する．2次元の画像から3次元情報を復元するshape-from-Xの問題など，多くのものはこの不良設定問題である．そこで何らかの知識を制約条件として加え，この問題を解くことになる．その際，多くの手法では場当たり的にその制約条件を加えているが，そのようなアプローチでは個別事例的で統一性がない．正則化手法[19]では，制約条件を関数の形で表現することにより，制約充足問題を関数の最適化問題という統一した考え方で解こうするアプローチである．そこには以下に示す2つの正則化手法がある．

(1) 決定的正則化

決定的正則化を利用する場合には，解の性質として区分的に滑らかな関数を考え，滑らかな部分における1次または2次の勾配がある範囲内に収まるという制約を加えることが多い．その際に，その範囲を越える部分が存在する場合には，これを不連続境界とする．この不連続境界をいかに取り扱うかがひとつの問題となる．それを解くための手法[19]として，スプライン曲線に重み関数を導入した手法や，GNC（graduated non-convexity）法などが提案されている．

(2) 確率的正則化

確立的正則化は，決定的正則化に比較して広い範囲の問題が扱える手法である．また不連続境界に対してはGemanが提案したline process[26]などが利用できる．しかし，解を確率的に探索していくために計算量が膨大になるという問題点がある．決定的正則化で解ける問題はそちらを利用するほうが効率的なので，今後は決定的正則化では解けない問題をこの手法で解くことなどがあげられる．

### 3.3 Physics-based vision

これまで直感にまかせて実現されてきたCVシステムを，より科学的に構成してゆこうとする動きのひとつにphysics-based visionがある．われわれが観測している

画像上の各点の明るさは，その点に対応する部分の物体の面の向き，物体表面の反射率，照明，周囲物体の表面での反射など物理光学的な拘束によって決定されている．その物理光学的な手法を積極的にビジョンの中に取り込もうとする方法が physics-based vision である．典型的な例としては，物体の表面の反射特性を利用して表面の形状を復元する shape-from-shading などがあげられる．最近では物体の表面の反射モデルとして，より正確な物理特性を表現する Torrance-Sparrow のモデルなどが利用されている[27]．これまで，光学原理を適用するには様々な仮定が必要となるために実用的なビジョンシステムへの適用性が低いという批判もあったが，physics-based vision を実用的なシステムに応用しようとするアプローチもみられる[28]．

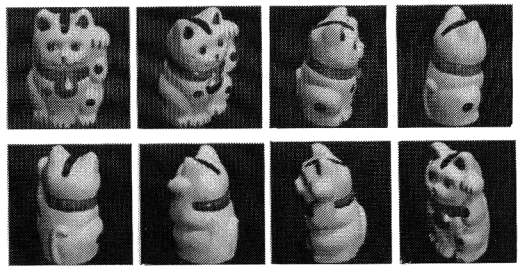
## 4. アピアランスベースの手法

特別な特徴抽出を行うことなく物体の見かけ画像をそのまま利用するアプローチが，アピアランスベースのアプローチである．エッジやコーナーなどの特徴を抽出する必要がないために適用範囲が広い．その手法と固有空間法とが結びついた手法が最近着目されている．固有空間を画像処理に応用する研究は古くからなされており，例えば統計的パターン認識[29]における主成分分析の利用や，画像の符号化などにみられる KL（Karhunen-Loève）展開などがよく知られている．しかし，従来画像のアピアランスのような高い次元のベクトルに対する固有ベクトルの計算には膨大な時間がかかり，実用的ではなかった．それが1章で述べたように，最近の計算機の進歩により小型ワークステーションでも計算できるようになった．さらに，対象を画像に限り，その性質を利用して固有ベクトルを高速に計算する手法[30]も提案されている．認識に応用した例として，MIT では主成分分析（固有顔）を利用した顔画像認識[31]などを提案している．

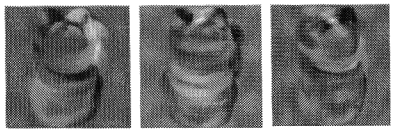
本章では，筆者らがアピアランスベースのアプローチとして研究を進めているパラメトリック固有空間法について簡単に紹介する．

### 4.1 パラメトリック固有空間法による物体認識[32]

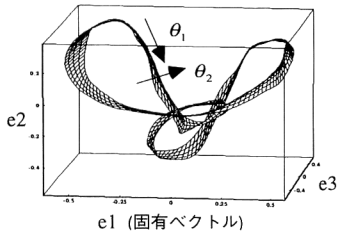
3次元物体の見かけの画像は，その物体の方向や照明の位置により大きく変動する．例えば，ある物体を1回転させただけで図2(a)に示すような多様な画像が得られる．最も単純なアピアランスベースの手法としては，すべての画像を単純に記憶し，入力画像と照合する手法が考えられるが，これでは膨大な記憶容量と計算時間が必要となる．多数の入力画像をいかに圧縮して記憶させるかがここでの学習の問題となる．



(a) 物体を回転させた時の見かけ画像の変化



(b) 固有ベクトル




(c) パラメトリック固有空間

図2　3次元物体の学習（パラメトリック固有空間法）．

### 4.2 物体の学習

物体の学習を2段階の処理から構成する．第1段階目では，学習に使用する多数の画像集合を主成分分析により圧縮する．学習に使用する画像は，学習すべき3次元物体の方向を変化させ，さらに光源位置も変化させながら収集する．収集された画像集合は膨大なデータとなるが，これらを主成分分析により少数の固有ベクトルの線形結合で表現できる．図2(b)に固有ベクトルの例を示す．

学習の第2段階目では，物体の方向や光源の位置の変化により連続的に変化する画像を固有空間上の多様体で表現する．固有空間とは固有ベクトルを軸とする空間であり，1枚の画像は固有空間上の1つの点に対応する．例えば，1回転分の学習サンプル系列を固有空間上に投影させると，投影点は1次元の点の系列になる．一般的にサンプル系列中の連続する2枚の画像間の相関は強いため，この点系列は固有空間上を滑らかに動くことになる．これは，相関が強い2枚の画像に対応する固有空間

上の2点は距離が近くなるという性質によるものである．連続する点の間はキュービックスプラインなどの適当な関数で補間する．

さらに，光源の位置を変化させた場合についても同様に固有空間上に投影させ補間処理を施すと，物体のポーズと光源の位置の2パラメーターで表現される多様体（曲面）が固有空間上に構成される．物体の種類が複数の場合にはその数だけ多様体が構成されることになる（図2(c)）．物体の方向を $\theta_1$，光源の方向を $\theta_2$ とすると，物体 $p$ に対する多様体は $\boldsymbol{g}^{(p)}(\theta_1, \theta_2)$ で表現できる．学習サンプルに存在しない方向や光源の位置（中間の方向や光源の位置）に対する画像も，補間によりこの多様体で表現されていることになる．また，物体の回転については，今回の例では1軸回りの回転，つまり回転に対して1自由度の場合を扱っているが，自由度が増えた場合にも多次元の多様体に拡張することにより容易に対処可能である．認識すべき物体のクラスの数を $p$ とすると，このような多様体を $p$ 種類，固有空間の中に作成することにより，物体の学習は終了する．

### 4.3 物体の認識

認識すべき物体を含む入力画像において，まず物体の領域を背景画像から切り出す．その後に大きさの正規化を行い，入力ベクトルを作る．次にこのベクトルを固有空間上の点 $\boldsymbol{z}$ に投影する．認識段階では，この投影された点 $\boldsymbol{z}$ がクラスの数 $p$ 種類ある多様体のどこに近いかを調べることになる．つまり，点 $\boldsymbol{z}$ と曲面 $\boldsymbol{g}^{(p)}(\theta_1, \theta_2)$ との距離

$$d_1^{(p)} = \min_{\theta_1, \theta_2} \| \boldsymbol{z} - \boldsymbol{g}^{(p)}(\theta_1, \theta_2) \|$$

を最小とする $p$ を求めることにより実現できる．その際，距離を最小とする $\theta_1$ が物体の方向，$\theta_2$ がそのときの光源の方向を表している．つまり物体の向きや光源位置を検出することができる．もし，この点 $\boldsymbol{z}$ がどの多様体にも近くない場合には，この入力画像は学習された物体ではないと判定される．

### 4.4 固有空間法の応用

上記で述べたパラメトリック固有空間法は汎用性の高い画像の表現法であるために，様々な応用が考えられ，その一例を以下に示す．

(1) 物体や顔の切り出し

パラメトリック固有空間の考え方を発展させ，物体の向きと物体の大きさの変動をパラメーターとし，画像に局所的にこの認識法を適用することにより，認識と切り出しとを結合した認識が可能となる．それを用いて，複雑な背景中から任意の方向を向いた任意の大きさの3次元物体を検出する手法[33]が提案されている．

(2) ビジュアルサーボ

上記したように，パラメトリック固有空間法により物体の向きや物体の位置が検出できる．これをマニピュレーターにフィードバックすれば，物体の位置決めなどに利用できる．これはビジュアルサーボ[34]などへの応用が可能である．

(3) 照明の解析

物体の見かけ画像は，照明の方向や照明の色により大きく変化する．物体認識にとって都合のよい照明とは，対象となる物体間で画像間の類似性が少なくなるような照明である．照明の変化による見かけ画像の変化をパラメトリック固有空間で表現することにより，この問題を効率よく解く手法が提案されている[35]．

(4) そ の 他

その他，固有空間法を陰影からの形状復元に利用する例，距離画像に利用する例，局所的な特徴に利用する例なども提案されている．これらについては文献21)を参照していただきたい．

本解説では，最近のコンピュータービジョン研究の動向，および筆者らが進めている固有空間法による画像認識手法を紹介した．最近の計算機の能力の飛躍的な向上および画像入出力機器の一般への普及という背景の中で，従来からの学際的なCV手法に加え，実環境で利用できる手法，また目的を実現するための直接的な手法の研究が着目されるようになってきている．

日頃ご指導いただくNTT基礎研究所，池上氏，石井氏，萩田氏に感謝する．

## 文 献

1) D. Marr：ビジョン―視覚の計算理論と脳内表現―（産業図書，1987）．
2) 辻 三郎："コンピュータビジョンへの道"，情報処理，**24**（1983）1414-1420．
3) 宇都宮敏男，ほか："特集「認識と理解」"，電子情報通信学会誌，**71**（1988）1111-1256．
4) 大田友一，杉原厚吉，ほか："小特集「コンピュータビジョンにおける手法」"，情報処理，**30**（1989）1026-1074．
5) 大橋昭南，絹川博之，中村英夫，ほか："小特集「医用画像と情報処理」"，情報処理，**30**（1989）206-224．
6) A. Rosenfeld and A. C. Kak（長尾 真ほか訳）：ディジタル画像処理（近代科学社，1976）．
7) D. H. Ballard and C. M. Brown: *Computer Vision* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1982).
8) B. K. P. Horn: *Robot Vision* (The MIT Press, Cambridge, 1986).

9) 白井良明：パターン理解（オーム社，1987）.
10) 金谷健一：画像理解（森北出版，1990）.
11) 谷内田正彦：ロボットビジョン（昭晃堂，1990）.
12) 松山隆司："AIマップ－ビジョン研究から見た統合アーキテクチャー"，人工知能学会，**10**（1995）888-894.
13) 田村秀行，池田克夫：知能情報メディア（総研出版，1995）.
14) V. V. Vinod, C. Hashizume and H. Murase: "Focussed color intersection for content-based retrieval," 情報処理研究会，CV-97-7（1995）45-52.
15) 金子正秀："知的画像情報処理"，テレビジョン学会，**49**（1995）750-759.
16) 和田俊和："コンピュータビジョンにおける技術評論と将来展望(Ⅰ)－投票と多数決原理に基づく幾何学的対象の検出と識別－"，情報処理研究会，CV-91-10（1994）71-78.
17) 杉本晃宏："コンピュータビジョンにおける技術評論と将来展望(II)－ビジョンにおける不変量とその応用－"，情報処理研究会，CV-93-3（1995）19-34.
18) 石黒　浩："コンピュータビジョンにおける技術評論と将来展望(III)－能動視覚－"，情報処理研究会，CV-94-6（1995）45-52.
19) 天野　晃："コンピュータビジョンにおける技術評論と将来展望(IV)－正則化－"，情報処理研究会，CV-95-9（1995）59-66.
20) 鶴田直之："コンピュータビジョンにおける技術評論と将来展望(V)－多層ニューラルネットワーク－"，情報処理研究会，CV-96-3（1995）17-24.
21) 村瀬　洋："コンピュータビジョンにおける技術評論と将来展望(VI)－固有空間法による画像認識－"，情報処理研究会，CV-97-9（1995）59-66.
22) 久野義徳："アクティブビジョン－歴史と展望－"，人工知能学会誌，**10**（1995）493-499.
23) R. Bajcsy: "Active perception," Proc. IEEE, **76**（1988）996-1005.
24) J. Aloimonos: "Purposive and qualitative active vision," *Proc. International Conference on Pattern Recognition*（1990）pp. 346-360.
25) D. H. Ballard: "Animate vision," Artif. Intell., **48**（1991）57-86.
26) S. Geman and D. Geman: "Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell., **6**（1985）721-741.
27) S. K. Nayar: "Shape recovery using physical models of reflection and interreflection," Ph. D. dissertation, Carnegie Mellon Univ.（1991）.
28) 和田俊和，浮田浩行，松山隆司："イメージスキャナを用いた書籍表面の3次元形状復元—近接照明下でのShape from shading"，信学論，**J77-DII**（1994）1059-1067.
29) K. Fukunaga: *Introduction to Statistical Pattern Recognition*（Academic Press, London, 1990）.
30) H. Murase and M. Lindenbaum: "Spatial temporal adaptive method for partial eigenstructure decomposition of large images," IEEE Trans. Image Processing, **5**（1995）620-629.
31) A. Pentland, B. Moghaddam and T. Starner: "View-based and modular eigenspace for face recognition," *Proc. Computer Vision and Pattern Recognition*（1994）pp. 84-91.
32) H. Murase and S. K. Nayar: "Visual learning and recognition of 3-D objects from appearance," Int. J. Comput. Vision, **14**（1995）5-24.
33) H. Murase and S. K. Nayar: "Image spotting of 3D objects using parametric eigenspace representation," *Proc. of Scandinavian Conference on Image Analysis 95*（1995）pp. 323-332.
34) S. K. Nayar, H. Murase and S. A. Nene: "Learning, and positioning, and tracking visual appearance," *Proc. IEEE Int. Conf. on Robotics and Automation*（1994）pp. 3237-3244.
35) H. Murase and S. K. Nayar: "Illumination planning for object recognition using parametric eigenspace," IEEE Trans. Pattern Anal. Mach. Intell., **16**（1994）1219-1227.