〔解 説〕

センサ統合とコンピュータビジョン

浅 田 稔*

1. はじめに

画像処理、画像理解など、人間の視覚能力を機械で実現させるための研究分野を、一般にコンピュータビジョンと呼ぶ。コンピュータビジョンの中心課題は、2次元の画像を処理して、元の3次元世界のモデルを作成し、その中にある物体の形や配置の記述を得ることである¹⁾。基本的な処理の流れを図1に示す。簡単に説明すると、シーン（3次元世界）を写した画像がシステムに送られ（画像入力）、それから明るさの急変する所などの画像上の特徴（2次元パターンの特徴）を抽出する。次に、これらの特徴を元に、3次元世界の特徴（例えば、距離や面の法線方向、辺の凹凸等）を復元する。最後に、前もってコンピュータに蓄積されている物体のモデルと照合し、どのような物体が、どの位置と向きにおかれているか、他の物体とどのような関係にあるかを知り、更に必要に応じて、隠れている部分を推定したり、シーン全体の意味を推論する。

これらの段階の中で、2次元画像上の特徴からシーンの3次元特徴を復元するには、基本的に情報が欠落しているため、なんらかの拘束条件が必要となる。これまで、対象とするシーンを限定し、弛緩法や正則化等を駆使して、3次元形状を推定する研究が盛んに行われてきた²⁾。コンピュータビジョンが元々、ロボットの視覚の役割を果たすことを考えると、ロボットが静止して、じっとシーンを見つめている状況は考えにくい。ロボット自身がタスクを遂行するために移動したり、またシーン中に動く物体などが存在し、視覚センサから得られる情報は、時々刻々と変化する。即ち、現実の世界では、なるべく少ない情報でもとの3次元世界の情報を再構成するのではなく、なるべく多くの情報を利用して、より正確に、効率よく、豊富な世界モデルを構築していかなければならない³⁾。

多くの情報は、多種のセンサ、多数の観測点を意味し、

これら进行处理する手法として「センサ統合」という言葉が用いられており、現在、盛んに研究されている。本稿では、センサ統合をコンピュータビジョンの分野から眺め、2でその目的・利点について、3で筆者らのグループの研究を紹介しながら、問題点、今後の課題などを探る。

2. コンピュータビジョンとセンサ統合

2.1 センサ統合の役割

センサ統合システム構築の難易度を決定する要素は、1) 与えられるタスクの難易度、2) 環境の複雑さ（屋内、屋外（道路シーン、クロスカントリー）、既知、未知）、3) 利用できるセンサの種類、性能等の3つである。これらのうちで、比較的明確なのは、3) のセンサに関する部分である。各センサについて、それが抽出する特徴、精度、特性などはほとんど既知であり、一般的な形としては、センサモデルとして記述される⁴⁾。

1) と 2) に関しては、明確な定義が困難である。即ち、タスクとしては、画像の領域分割、物体認識、シーンの解釈から、自律移動ロボットの応用として、道路追従、障害物の検出及び回避、ある特定の物体（ランドマーク等）の発見などが考えられる。また、ある時には、「次の交差点を右折して、すぐ右にある小屋の前で止まりなさい」などの形で与えられることもあり、種々のレベルを考慮しなければならない。これらのことは、タスクプランニングや行動計画の問題としても研究されている。2) の環境の複雑さは、与えられるタスクにも依存する。例えば、複雑そうに見える自然環境でもレンジファインダを用いた障害物の検出だけであれば、そんなに困難で

図1 コンピュータビジョンの基本的な枠組み

原稿受付 1989 年 10 月 4 日

Integration of Multisensory Information and Computer Vision

* 大阪大学工学部電子制御機械工学科

図 2 道路シーンから得られた
視覚情報

図 3 高さ地図の鳥瞰図

図 4 分割された高さ地図

はない。また、箱だけが置かれた倉庫のような環境でもある特定の箱をそのサイズを正確に計測することによってしか発見できない場合は、そんなに簡単ではない。

ロボットに要求されるタスクが、高度になればなるほど、人間と同等の環境理解能力が必要となり、ビジョンの果たす役割は大きい。即ち、TV カメラやレンジファインダなどを通して得られる視覚情報は、空間的に配置された明度や距離の観測値の単なる配列ではなく、それ自体で、センサ情報の集合であり、それらの相互関係に

より、全体として、より多くの情報量を持ち、シーンの大局的な情報を与える。これが他のセンサ情報の処理システムと大きく異なる点である。そのため、視覚情報の処理や理解だけで信号処理から意味処理までの多様な処理レベルを含み、一つの大きなシステムになる。よって、コンピュータビジョンの研究分野では、センサ統合は、それ自体が目的では無く、最終目標である認識、理解、解釈などのより上位レベルのタスクを遂行するための手段として扱われる事が多い。

2.2 センサ統合の利点

多種、多視点のセンサ情報を利用するセンサ統合の利点として、以下の事柄が挙げられる。

1) 冗長性：異なるセンサでも環境に対する同じ特徴を抽出する場合、もしくは、同一センサでも複数の観測点から同じ特徴を抽出する場合、冗長な情報が得られる。これらを融合（統合）することにより、センサ情報の不確実性を減少させ、精度向上が計られる。また、センサ情報に対する信頼性も向上する。センサ出力の物理的な次元が同じなので扱い易く、システムの下位のレベルで数値的に融合される。これまでの多くのセンサ統合手法で実現されている。

2) 相補性: 異なるセンサから得られる特徴が互いに独立 (異なる特徴) しており, 個別では, 達成できないが, 同時に利用することによって初めて達成可能な場合を指す. 例えば, カラーカメラから得られる画像からは色やテクスチャに関する情報が得られ, レーザレンジファインダからは3次元の情報が得られる. これらを同時に利用することによって, 物体の形状と色情報が得られ, 物体の同定が容易になる. また, ある場合には, 融合されることなく, あるセンサ出力がそのまま適用されることにより, 別のセンサの探索範囲を限定したり, 制御することも含まれる.

3. センサ統合による道路シーンの解釈と環境モデルの構築

以下では, センサ統合を利用したコンピュータビジョンの研究例として, 明度画像と距離画像を用いた道路シーンの解釈と環境モデル構築法について紹介する^{5)~7)}. 本システムの最終目標は, シーンを解釈し, 環境モデルを構築することであり, センサ情報の初期処理から, 認識・理解などの処理のレベルの各段階で, センサ統合的手法を利用している. ここでは, 処理のレベルと利用されているセンサ統合の利点を考慮して, 3つのトピックを紹介する. その前に, 使用したセンサ情報とシステム的前提条件, 及び, 3つのトピックに共通の前処理を示す.

3.1 入力シーンと前提条件

図2(a), (b)に示すのは, 移動ロボットの視覚装置から得られた道路シーンに対する, ある時刻の明度画像とそれに対応する距離画像である. 入力されたシーンは中央に道路があり, その道路上にはトラック, 両側には小屋やポスト, 垣根や木などがある. シーン中に2つの動物体がある. 車は, 交差している道路を右から左へ通過する. 距離画像では, 明るい部分が観測者からの距離が近い点で, 暗い部分が遠方の点を示している. 黒い部分は, 距離情報が得られなかった部分で, 反射が不十分であったり, 他の物体に隠されている部分 (隠ぺい領域) に相当する.

本システムでは, 道路シーンを対象とするが, 以下の事柄を前提条件としている.

(1) 道路は, ほぼ平坦な地面上に存在するとし, クロスカントリーなどの急激な地面の凹凸は無いものとする. この条件により, シーンは, 大きな水平面とその上にある物体によって構成されていると考えることが出来る.

(2) 道路上やその近傍にある物体としては, 人工物として自動車, 建物, 道路標識, 郵便ポスト, 自然物として高木, 灌木を対象とし, 3.5で紹介するシステムで

は, それらに関する一般的な知識を前もって与える. これら以外にも, 人間, 自転車など数多くのものが道路シーンでは考えられるが, それらについては, 解釈せず (知識として与えず), 未知物体として取り扱う.

(3) センサ情報として, ある時刻に得られたTVカメラからの白黒濃淡画像とレンジファインダから得られる距離画像を扱う. TVカメラとレンジファインダのそれぞれの幾何学パラメータや相互の幾何学的位置関係も既知とする.

3.2 前処理

図2(a)に示すような複雑なシーンを明度画像のみを用いて, 解釈することは, 困難であり, 距離情報との併用により, 容易になると期待できる. 距離画像は, 観測者からシーン中の物体の各点までの距離を記録したものであるが, 距離画像自体は, センサ (レンジファインダ) 中心の透視変換によって構成されているので, 距離画像から, 直接, 元の3次元世界のレイアウト等を推定することは容易ではない. そこで, 前提条件の(1)と(3)を考慮して, 距離画像に記録されている距離情報を, ロボットが存在すると想定される地面を座標平面としたカーテジアン座標系の地図上の高さ情報に変換する. 図3にその鳥瞰図を示す. 小屋の屋根の一部やトラックが存在することが窺われる. 高さ地図では, 地面に垂直な方向の距離情報は1点に集約されるので, 垂直な面を検出できない. そこで, 高さ地図の各点に対して, その点の最高及び最低の高さを記録した多重高さ地図を作成する. これにより, 垂直面の検出が可能になる. 図3では, 最大高さを示している.

文献⁶⁾では, 実際のセンサ中心の座標系で表現された地図を「実センサ地図」と呼んでいる. 図2(a, b)等は, 実センサ地図の例である. また, 実センサ地図から, センサ情報を変換して得られる地図のことを「仮想センサ地図」と呼び, 高さ地図も「仮想センサ地図」の一種としてとらえている. 仮想センサ地図上のデータは, 実センサ地図からの情報を座標変換して表現しているに過ぎないので, 量的には情報は増えていないが, 視点を変えてセンサ情報を解釈するという仮想センサ地図の考え方により, 異なったセンサ情報を表現したり, また3.4で示すように, それらをカーテジアン座標系で統合することが容易になる.

前提条件(1)を考慮すると, 高さ地図は, ロボットが移動可能な地面とその上の物体に分割できる. 先ず, 最初に, 高さ地図は, 2つのタイプの領域に分けられる. 高さ情報が得られる領域とそうでない領域である. 高さ情報が得られない領域は, 視野外領域か隠ぺい領域であると考えられる. このうち視野外領域は, レンジファイ

図 5 高さ地図の利用による明度画像の分割

ンダのキャリブレーションパラメータ（高さ、傾き、走査する角度）から簡単に見つけることができる。残った領域が隠れ領域となるが、垣根や木といった自然物は、表面の凹凸が激しく反射が十分に得られないので、高さ情報が得られないことがよくあり、隠れ領域に分類されやすい。高さ情報が得られる領域は、移動可能領域か障害物領域である。移動可能領域の発見は容易である。最初に、予想した地面に非常に近い点を集め、原始領域とする。次に、その周囲にある傾斜や曲率の小さい点を次々に併合することによって原始領域を拡張させ、移動可能領域とする。高さ情報が得られている領域で、移動可能領域以外が障害物領域である。図 4 に分割された高さ地図を示す。黒い部分、暗い灰色の部分、明るい灰色の部分、白い部分がそれぞれ視野外領域、移動可能領域、隠れ領域、障害物領域に相当する。

以下では、明度画像、距離画像、高さ地図などのマルチセンサ情報を駆使した処理例を紹介する。

3.3 距離情報と明度情報による領域分割

TV カメラなどから得られる明度画像の領域分割問題は、シーンを解釈する上で、非常に重要な問題であり、

平面近似可能領域
距離データ分割
明度データ分割

図 6 距離情報と明度情報を利用した領域分割

古くから、コンピュータビジョンの分野で研究されてきた⁸⁾。一般に、シーンが複雑になると、明度情報だけでは、意味ある領域に分割することは、困難である。明度情報に比べて、距離情報は、安定した幾何学的情報を与える。よって、距離情報と明度情報を併用することによって、複雑な屋外シーンの明度画像の分割を簡単化することが期待できる。

図 2 に示したシーンの解析⁵⁾では、図 4 に示した障害物領域を解釈する時、高さ情報だけでなく、明度情報を利用するために、明度画像を分割した。その方法は、高さ地図上の障害物の輪郭線を明度画像に投影する事である。高さ地図上の障害物の輪郭部を明度画像に射影することは、レンジファインダとカメラの位置関係が既知なので簡単にみえる。しかしながら、各点の高さ情報の最

図 7 6 枚の高さ地図の統合

高値と最低値を選択しなければ、明度画像内の障害物領域をうまく分割できない。例えば平面上に立方体があると仮定し、斜め上方から TV カメラで観測したとしよう。高さ地図から最大の高さ情報だけを使って射影すると、TV カメラから見た画像では上面しか検出されず、側面は発見されないままである。そこで、高さ地図上の障害物領域の輪郭線上の各点について、その点が他の物体を隠している点か、他の物体によって隠されている点かを識別し、射影する高さ情報（最高もしくは最低）を多重高さ地図から選択する。

図5に上記の方法によって明度画像に投影された障害物領域を示す。トラックや、右側の小屋の手前にある車、郵便ポスト、停止標識、垣根など正確に射影されているのがよくわかる。交差点内の自動車は、観測者から遠方にあり（高さ地図の範囲外）、投影されていない。左側の小屋の屋根の境界は、雑音の多い高さ情報のため正しくない。右側の小屋の屋根の上部の境界が、途中で急激に落下しているのは、小屋のうしろに低い物体があり、障害物領域としては小屋とこの物体を含んでいるためである。

この分割では、一つの障害物が明度画像上で一つの閉領域に対応し、その内部は、分割できない。それは、高さ地図からは、高さ情報しか利用できないからである。これに対し、距離画像を明度画像に射影し、両方の情報を利用した明度画像の領域分割も考えられる⁹⁾。距離情報が利用できる部分に対して、先ず平面近似可能領域を抽出する。平面近似誤差が大きい部分は距離の連続性を用いて領域分割する。次に距離情報が利用できない部分に関しては、明度の分散を用いて領域分割する。図6に、分割結果を示す。

この他にも、明度情報（カラー画像）を距離画像上に射影してカラー距離画像を作成し、カラー情報と距離情報を相補的に利用して、階段の検出に応用した例や¹⁰⁾、明度画像から得られた道路境界を距離画像に射影して、障害物と道路境界の位置関係から、操舵角を決定した例¹¹⁾もある。

3.4 時系列距離画像の統合⁹⁾

2次元の画像から抽出された3次元情報を高精度化するために、多視点観測がよく用いられている。以下では、多視点観測によるセンサ情報という観点から、時系列距離画像の統合について紹介する。

移動ロボットの視覚では、レーザレンジファインダ等を用いて、直接、観測者から、シーンの各点までの距離を表す距離画像を得ることができる。多視点から得られた距離画像を統合することは、既に、3次元情報が得られているので、容易に見えるが、実際には、多視点で観

(a) 明度画像上の移動可能領域と道路境界の候補

(b) 高さ地図上の道路境界

(c) 推定された道路領域

図8 道路領域の抽出過程

測されたデータは、同じ物体に対しても、(1) 分解能、精度が異なり、それらをどの様に統合するか、また、(2) 矛盾するデータをどの様に処理するか等の問題を解決しなければならない。

(1)の問題をもう少し詳しく説明する。通常、距離画像は、レンジファインダ中心の座標系から透視変換を受けて構成されるので、画面間で距離情報を統合する際に

図9 トラックの領域に対するモデリングの過程

(a)

(b)

図10 解釈の結果(a)と環境地図(b)

は、3.2 で示したように、地面を座標平面とするカーテジアン座標系に距離情報を射影し高さ地図に変換する。しかしながら、距離画像と高さ地図では、分解能が異なるので、元のセンサ情報を正確に表現することは難しい。この問題に対し、我々は、元のセンサ情報の分解能や精度及び、異なる視点間での高さ情報の無矛盾性を元にして、高さ地図上の各点の高さ情報の信頼性を定義し、これを用いて、高さ地図の照合、更新を可能にしている⁹⁾。これまでの時系列高さ地図の統合の研究¹²⁾などと異なる点は、動物体を含むシーンを扱っているところである。動物体を含まない場合には、対象シーンは一つの剛体であり、シーン全体に統計的な手法が適用できるが、動物体を含む場合には、容易ではない。定義された高さ情報

の信頼性を利用することにより、シーン中の動物体の発見、矛盾するデータの取扱を容易にした。図7に6枚の高さ地図を統合した結果を示す。道路シーンは図3と同じである。道路上のトラックの追跡が成功し、位置が更新されている(直進後右折)様子がわかる。

3.5 距離画像と明度画像の統合による屋外シーンの解釈

複雑な道路シーンを解析し、個々の物体を自動車や小屋などに解釈するためには、シーンに対する知識が必要である。文献⁷⁾では、シーンに対する一般的な知識をフレーム構造で表現し、距離画像と明度画像から、屋外シーンの解釈と環境モデル構築を動的に制御する手法を提案した。3.3 で示した処理結果(距離画像を高さ地図に変換し、その分割)をガイドとして、障害物に対する解釈をセンサ情報やそれまでの解析状況に応じて動的に制御する。ここでは、物体間の関係(例えば、自動車は道路の上を移動する)を利用して、仮説・検証、フィードバックを繰り返しながら信頼性の高いシーンの解釈を可能にしている。また、解釈のレベルに応じた物体の幾何モデルを作成し、これにより、観測されていない未知の部分に対する幾何学的な推論を可能にしている。

図8に、道路領域に対する過程を示す。高さ地図の解析から得られる移動可能領域は、道路領域を含むが、道路境界は明度画像の処理により検出可能である。そこで、高さ地図上の移動可能領域を明度画像に射影し、探索領域を限定する(図8(a)参照)。次に得られた道路境界を元の高さ地図上に射影(図8(b)参照)し、道路幅一定という知識から、道路領域を推定する(図8(c)参照)。

この推定により、観測事実からは、道路の傍にあるとされた人工物(実はトラック)が、道路上にあると修正され、自動車であると仮定される。自動車のモデルとして、トラックと乗用車があり、サイズなどから、トラックと推定される。これらの解釈に応じて、物体の幾何モデルが作成され、センサ情報を利用して、細部が決定されていく。図9に、トラックに対するモデリングの過程を示す。また、図10(a)、(b)は、最初の観測点で得られたセンサ情報の解釈結果である。図中、人工物(A)、自然物(N)は、それぞれ、直方体、円筒で近似されている。また、トラックや小屋などは、それらの解釈に応じて、多面体で表現されている。

多種センサ情報の統合利用により、道路境界の抽出にみられたように、探索領域が限定され、効率的かつ正しくシーンの解釈が行えた。またトラックの同定プロセスで示したように、解釈の途中結果が幾何モデリングに反映され、高精度化された幾何パラメータがシーンの解釈

を効率的かつ信頼性の高いものになっている。

☆ ☆

センサ統合とコンピュータビジョンとの関係を筆者らの研究事例を紹介しながら明らかにした。コンピュータビジョンの研究は、先にも述べたように、センサ情報の信号処理から、認識・理解などの上位レベルの意味処理も含み、広範に渡る。このため、複数のセンサ情報の利用や、多視点観測による統合は、コンピュータビジョンの各レベルでのゴール達成のための強力なツールとなっている。まとめと課題として、以下の事が挙げられる。

(1) シーンの解釈に対する信頼度の表現は、多種多視点の観測データを統合する上で重要な問題である。物理的な次元が同じ場合、確率を用いて、表現、更新している例があるが¹³⁾など、物体のより高度な解釈に関してその信頼度を明確に定義することは難しい。3.5で紹介したシステムでも、各段階で最も信頼性のある特徴をヒューリスティクスに基づいて選択する。一般的には、各特徴とその評価ルーチンにしたがって信頼度を計算し、その値にしたがって解釈に利用できる特徴を決定すべきであり、そのような信頼度の計算法を確立する必要がある。

(2) 3.5で紹介したシステムでは、物体の幾何モデルを利用して、空間推論を必要に応じて行っている。例えば、道路領域と障害物の位置関係を求めるとき、高さ地図上では、単に隣接しているに過ぎないが、道路幅一定という属性と隠ぺい領域に対する推論から、この物体が道路の「上」にあると推定している。この様な推論を未探索領域や、隠ぺい領域に対してより積極的に行うことが出来れば、環境モデルはより高精度化されるであろう。

(3) 多視点観測によるセンサ統合は、データの信頼性を増すと共に、新しい観測データの付加により、シーンの解釈・空間推論の検証及び修正に役立つ。特に、自動車などの移動物体は、画面間での対応づけにより、自動車であることの確認と、その運動パラメータが求められる。3.4では、「高さ」を物理的に共通な次元として統合しており、3.5で提案しているような高度なシーンの解釈結果は利用していない。多視点での観測データも含めて、より高度で柔軟なダイナミックシーンの解釈を行うことが考えられる。

謝 辞

本稿を始め、日頃から、ご討論頂いている本学工学部白井良明教授、研究室諸氏に感謝する。

参 考 文 献

- 1) 白井良明, 「コンピュータビジョン」, 昭晃堂, 1980.
- 2) 坂上, 横矢, 「弛緩法と正則化」, 情報処理, 30, 9, pp. 1047-1057, 1989.
- 3) R. Jain: Dynamic vision, Proc. of 9th Int. Conf. on Pattern Recognition, pp. 226-235, 1988.
- 4) H. F. Durrant-Whyte: Sensor Models and Multi-sensor Integration, Int. J. Robot. Res., Vol. 7, No. 6, pp. 97-113, 1988.
- 5) 浅田, 「センサ情報の統合と理解による移動ロボットのための世界モデルの構築」, 日本ロボット学会誌, Vol. 8, No. 2, pp. 160-170, 1990.
- 6) M. Asada, M. Kimura, and Y. Shirai, "Dynamic Integration of Height Maps into a 3-D World Representation from Range Image Sequences", Proc. of 3rd Int. Conf. on Computer Vision, 1990.
- 7) 浅田, 白井, 「マルチセンサ情報を動的に統合することによる道路シーンの解釈とモデリング」, 情報処理学会論文誌, Vol. 31, No. 12, 1990.
- 8) C. K. Kohl, A. R. Hanson, and E. M. Riseman, "Goal directed control of low-level processes for image interpretation", In Proc. of IU Workshop DARPA, Vol. 2, pp. 538-551, 1987.
- 9) 門野, 浅田, 白井, 「センサ統合によるシーンの解釈と環境モデリング」, 1990年度人工知能学会全国大会予稿集, 1990.
- 10) Y. Goto, K. Matsuzaki, I. Kwen, and T. Obatake, "CMU sidewalk navigation system: A blackboard-based outdoor navigation system using sensor fusion with colored-range images", Proc. Fall Joint Computer Conference, pp. 105-113, 1986.
- 11) R. T. Dunlay, "Obstacle avoidance perception processing for the autonomous land vehicle", Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 912-917, 1988.
- 12) I. S. Kweon and T. Kanade, "High Resolution Terrain Map from Multiple Sensor Data", Proc. of IEEE Int. Workshop on Intelligent Robots and Systems '90, pp. 127-134.
- 13) A. Elfes: Using occupancy grids for mobile robot perception and navigation, Computer Vol. 22, No. 6, pp. 46-57, 1989.

浅 田 稔

(Minoru ASADA)

1953年10月1日生。'82年大阪大学大学院基礎工学研究科制御工学分野後期課程修了。同年同基礎工学部制御工学科助手。'88年同工学部電子制御機械工学科講師。'89年同助教となり現在に至る。その間'86~'87年米国メリーランド大学客員研究員。コンピュータ・ビジョン、人工知能、ロボティクスの研究に従事。工学博士。情報処理学会'89年度研究賞受賞。電子情報通信学会、情報処理学会、日本機械学会、IEEE各会員。

(日本ロボット学会正会員)