コンピュータビジョン　64－　8
（1990.1.26）

センサ統合と環境モデルの構築

浅田稔

大阪大学工学部電子制御機械工学科

あらまし　多種、多視点のセンサ情報を処理する過程は「センサ統合（Sensor Integration）」と呼ばれ、現在、盛んに研究されている。本稿では、センサ統合とコンピュータビジョンの関わりを、統合の結果得られるべき「環境モデル」の構築という視点から、明らかにすることを試みる。先ず最初に、センサ統合をビジョン研究に適用する場合の問題点を明らかにする。次に、センサ統合の一般的な枠組を示し、これまでのセンサ統合の各手法について概観する。統合結果得られる環境モデルの実例として、自律移動ロボットのケースを紹介し、最後に、今後の課題を述べてまとめる。

Intergration of Multisensory Information and Construction of a World Model in Mobile Robot Systems

Minoru Asada

Dept. of Mech. Eng. for Computer-Controlled Machinery

Suita, Osaka 565, Japan

Abstract Interest has been growing in the use of multiple sensors to increase the capabilities of intelligent mobile robot systems. This paper describes the relationship between the computer vision researches and the multisensor integration in the context of how a world model is constructed. First, the problems in applying the methods of multisensor integration to the computer vision problems are pointed out. Next, a general view of the multisensor integration which involves the advantages, the approaches, and the general fusion methods of the multisensor integration is given. Finally, several examples of constructing a world model for a mobile robot are shown, and the current problems to increase the capabilities of intelligent mobile robot systems are suggested.

1. はじめに

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]。基本的に情報が欠落しているので、それを補うために、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これまで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く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、これらを処理する手法として「センサ統合」とう言葉が用いられており、現在、ロボティクスの分野で盛んに研究されている。本稿では、センサ統合とコンピュータビジョンの関わりを、統合の結果得られるべき「環境モデル」の構築という視点から、明らかにすることを試みる。以下では、先ず最初に、センサ統合をビジョン研究に適用する場合の問題点を明らかにする。次に、センサ統合の一般的な枠組を示し、これまでのセンサ統合の各手法について概観する。統合結果得られる環境モデルの実例として、自律移動ロボットのケースを紹介し、最後に、今後の課題を述べてまとめる。

2. センサ統合とコンピュータビジョン

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、道路追従、障害物の検出及び回避等の一般的なものから、ある特定の物体の発見や、「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。このことは、タスクプラ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ーを用いた障害物の検出だけであれば、そんなに困難ではない。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依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、全体として、より多くの情報量を持つ。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。そのため、これまでの研究では、処理内容を低減して、簡単なパターン検出や計測に限り、冗長なセンサ情報の高精度化を計ったものが多い。環境理解を含めて、高度なレベルまで処理する場合には、ビジョンは不可欠である。

センサ統合システムにビジョンが含まれる場合、もしくは、ビジョン研究でセンサ統合的な手法を適用する場合、次の問題点が挙げられる。1) 異なるセンサ情報をどの様に表現するか、2) それらをどの様にして獲得するか、の2点である。これらの問題は基本的にビジョンの問題でもあるが、センサ統合システムではより重要なポイントとなる。自律移動ロボットによる環境モデル構築が典型的な例として挙げられる。そこでは、多種センサ情報を環境モデルとして、どの様に表現、更新するか、それらをどの様に制御するかが扱われている。4. で具体例を示すが、その前に 3. でセンサ統合の一般的な枠組を述べる。

3. センサ統合の一般的な枠組

3.1 センサ統合の定義

多種、多視点のセンサ情報を処理する過程が「センサ統合 (Sensor Integration)」といわれているが、この他にも「センサ融合 (Sensor Fusion)」等の用語も多く用いられている。石川 [5] は、心理学や生理学で使用されている用語を考慮して、「センサ複合」、「センサ融合」、「センサ統合」、「センサ連合」をそれぞれを区別、定義し、まとめて「センサヒュージョン」と呼んだ。ここでは工学的な立場から、Luo ら [6] の定義を用いる。即ち、センサ統合とは、システムのタスク遂行を助けるために、多種のセンサから得られる情報を互いに作用しあうように用いることを意味する。センサ融合は、統合プロセスの各段階を意味し、より狭義に定義され、センサ統合と区別されている。しかし、これらの定義も厳密ではなく、各様相を示していると考えた方がよい。

3.2 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

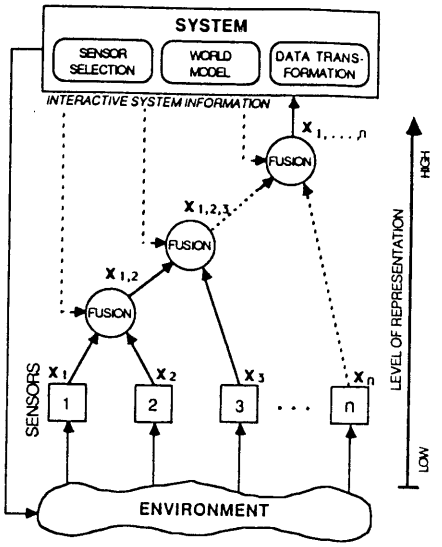


図1. センサ統合の一般的な形式 (文献 [6])

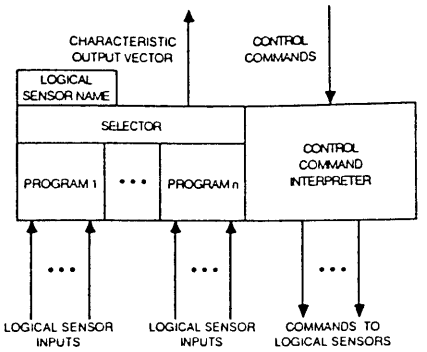


図2. 論理センサの基本モジュール (文献 [7])

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依って初めて達成可能な場合を指す。これは、シンボリックレベルでの融合と考えることも出来る。例えば、カラーカメラから得られる画像からは色やテキスチャに関する情報が得られ、レーザーレンジファインダーからは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、融合される事なく、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。

これらの他に、多種のセンサを用いたシステムの方が単一のセンサよりも、より速く、低いコストで実現できれば、センサ統合の利点と考えられる。

3.3 センサ統合の一般的なアプローチ

図1にセンサ統合の一般的な形式を示す。図中、丸印で示した部分で融合が行われ、全体で統合システムの過程を表している。全てのセンサ出力がネットワーク構造の中で統合されていく。システムを構成する主な要素は、現在の状況にしたがって適当なセンサ群をえらぶ「センサ選択 (Sensor Selection)」の制御戦略、センサ情報を表現する「世界モデル (World Model)」、そして融合の際や、世界モデルで表現する際にセンサ情報を変換する「データ変換 (Data Transformation)」等の知識を蓄えている知識ベースからなる。右端にあるスケールは、融合されたデータのレベルを示し、上位にいくほど、センサ情報が抽象化される事を意味する。

センサ統合の実際の構造としては、種々の形態があるが、ここではその代表として、「論理センサ」の考え方を紹介する。Henderson and Shilcrat[7] は、センサ統合の統一的な枠組を考える上で、実際のセンサをより抽象化した形で定義する「論理センサ」を提案した。論理センサでは、ユーザーは、実際のセンサの細かな処理に触れる必要はなく、センサの削除・追加が容易になる。図2に論理センサの基本モジュールを示す。このモジュールのネットワーク構造により全体のシステムを構成する。論理センサは、下位のいくつかの論理センサからの出力を受けて、新たな特徴量を決定する部分と、論理センサ間のインタフェース及びその管理を行う部分からなる。応用例として、レンジファインダーや消化器発見の論理センサネットワークなどが提案されている [7,8]。

3.4 センサ統合の各手法

これまで提案されてきたセンサ統合の各手法について紹介する。それらは、主に複数の冗長なセンサ情報を、統計的に処理して、精度を高める手法に基づいている。

1) 重み付き平均: 最も単純で、直感的な方法は、物理的に同一の実体を多視点で観測したセンサ情報を、何等化の方法で定義された重み付により、平均化する手法である。処理時間が短く、実時間性の実現も容易であるが、2) のカルマンフィルタの方が、統計的な意味で最適であることから、好まれている。移動ロボットシステム HILARE[9] で、3次元位置推定に利用されている。

2) カルマンフィルタ: システムが線形モデルで記述され、システムとセンサの誤差が、ガウスノイズで近似できるとき、カルマンフィルタは、融合データの統計的に最適な推定値を求めてくれる [10]。再帰的な性質と、低い計算コストから、多くのシステムで利用されている [11,12 など]。システムが線形で近似できない場合には、「拡張カルマンフィルタ」が利用されている [13]。

3) ベイズ推定 (Bayesian Estimate): センサ情報を確率分布で表現し、ベイズの定理にしたがって、冗長なセンサ情報を統合、更新する手法。4.1 で、その具体例 [14] を示す。

4) その他: 各センサがそれぞれ異なるベイズ推定を行うものとして、融合データを推定する多重ベイズ推定 [4] や、統計的決定理論 (Statistical Decision Theory) を応用した手法も提案されている [15]。

4. 環境モデル構築の例

ビジョンを用いたセンサ統合の例として、自律移動ロボットにおける環境モデル構築システムを紹介する。それぞれのシステムは、多少目的や条件が異なるが、多種、多視点の観測データを融合し、環境の表現を得る事を目標にしている。

4.1 占有格子を用いたソナーと一次元ステレオの融合

Elfes[14] は、移動ロボットにリング上に設置されたソナーからのセンサ情報と一走査線のステレオによる距離情報をベイズの確率モデルを利用して、融合・更新する世界モデル構築法を提案した。環境の地図として、占有格子と呼ばれる 2 次元格子を考え、各格子に対して、その場所が占有されているか自由空間であるかの確率をセンサ情報によって決定していく。一つのソナーからの距離情報による空間の占有確率密度分布を、図 3 のようなガウシャンセンサモデルを用いて近似する。この図では、左上方にソナーがあるとして、ピークの所までの距離情報が得られたときの分布を示している。ピークの所で物体が占有している確率が高く、その直前で最も低い (自由空間である確率が高い)。同様に、両眼立体視で得られる距離情報についても、視差に含まれる誤差から、物体までの距離推定について、確率分布が得られる。そこで、リング上に設置された 24 このセンサ情報と、一次元のステレオ視から得られる距離情報を、空間的に、また移動して得られる時系列の両方の情報を時間的に、ベイズの定理を利用して融合・更新した。

いま、k 番目のセンサの読みがR_kの時に、各格子が占有 (OCC) されている条件付き確率$P(OCC|R_k)$、(但し、自由空間 (EMP) である条件付き確率を$P(EMP|R_k)$とすれば、$P(OCC|R) + P(EMP|R) = 1$) が、既に更新されているとき、(k+1) 番目のセンサの読みR_{k+1}によってどの様に占有確率$P(OCC|R_{k+1})$が更新されるかを次式で示す。

$$P(OCC|R_{k+1}) = \frac{P(R_{k+1}|OCC)P(OCC|R_k)}{P(R_{k+1}|OCC)P(OCC|R_k) + P(R_{k+1}|EMP)P(EMP|R_k)}$$

占有されているときのセンサの読みが R である条件付き確率$P(R|OCC)$は、図 3 に示すガウシャンセンサモデルによる確率分布によって決定されるが、一般には、その決定が困難である。

図 4 に移動して得られた障害物地図の変化を示す。上段 ((a),(c), (e)) が、初期状態を、下段 ((b),(d),(e)) が、10 回目の更新時を表す。ソナー情報、ステレオ情報、それらの統合値をそれぞれ、左、中央、右に示す。各格子の濃淡が、空間が占有されている確率の高低を表し、＋印は、確率が 0.5(初期状態) であることを示す。

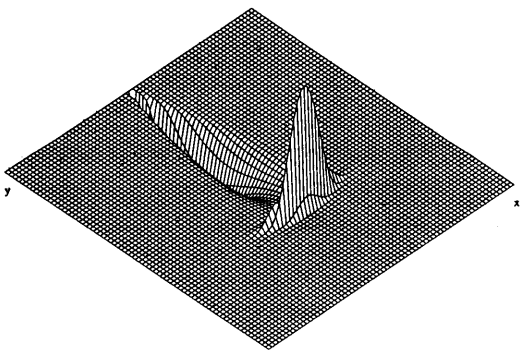


図 3. ガウシャンセンサモデルの確率密度分布 (文献 [14])

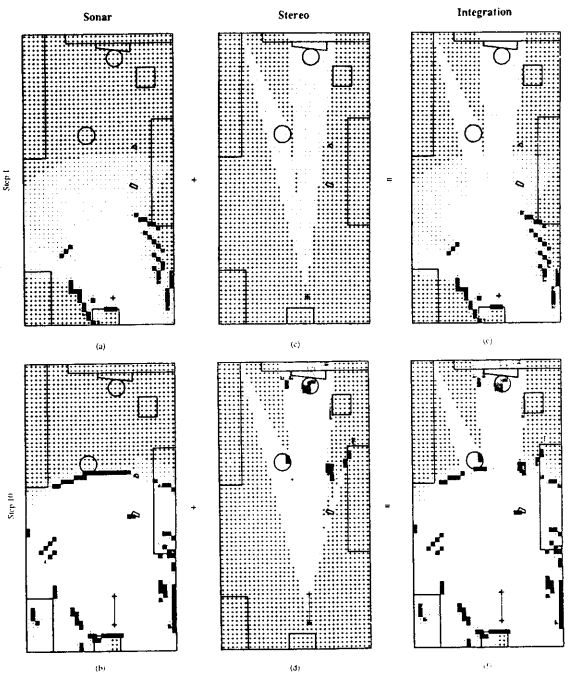


図 4. ソナーと一次元ステレオの統合結果 (文献 [14])

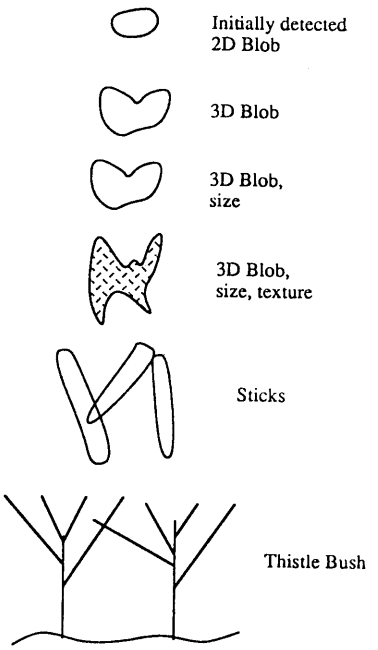


図 5(a). 表現空間による潅木の場合の多重表現 (文献 [16])

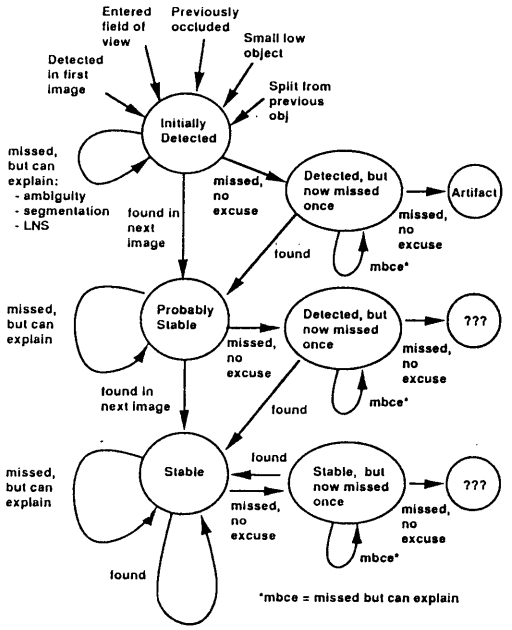


図 5(b). 潅木の場合の有限状態遷移図 (文献 [16])

4.2 表現空間を用いた物体の多重表現

上で述べた例では、各点に対して、占有されている確率を定めるので、空間表現は一種類しかない。また近傍の状態 (確率) が直接に格子の確率に影響を与えず、占有された格子の集合としての障害物の表現は存在しない。これに対し、Bobick ら [16] は、屋外の自然環境を対象として、移動ロボットから時々刻々得られる距離データの時系列画像から、センサ情報の信頼性や、状況に応じて物体の表現を変化させる方式を提案した。

最初に遠方で発見された物体は、信頼性が低いので、詳細に記述せず、単なる位置と大きさ程度の表現にとどめておく。観測者の移動により、徐々に近くなれば、それに応じて、記述を高精度化したり、新しい属性を加えて更新して行く。更に正確で詳細な情報が得られれば、それらを利用して、物体のクラスに応じた表現法を採用する。図 5(a) に潅木の場合の例を示す。

問題となるのは、いつ、どの様な状態で、表現を変化させるかであるが、各物体のクラス (ここでは、潅木) に対して、有限状態遷移図 (図 5(b) 参照) を用意し、これを利用して表現の変化を制御している。物体の多重表現により、より知的なナビゲーションが可能と主張している。

4.3 距離画像と明度画像の統合による屋外シーンの解釈

Asada ら [17、19] 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。距離画像を高さ地図に変換し [18]、これをガイドとして、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。先の例では、センサ情報の質に依存して物体の表現を変化させるだけであったが、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) も利用して、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、幾何学的な推論を可能にしている。図 6、7 に、物体に関する知識は表したフレーム構造のネットワークと幾何モデルの例をを示す。

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

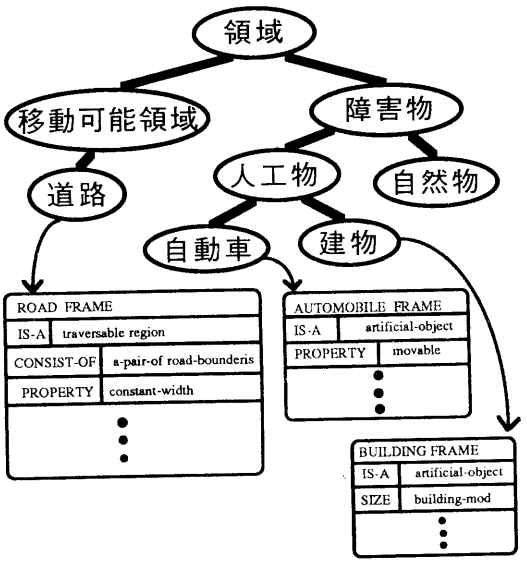


図6. フレーム構造を利用した物体に関する知識の表現

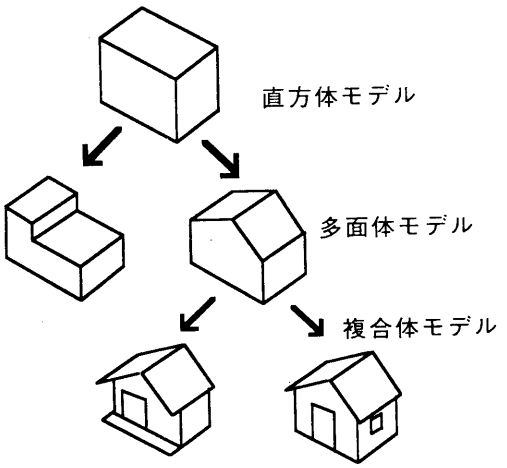


図7. 物体の幾何モデルの例

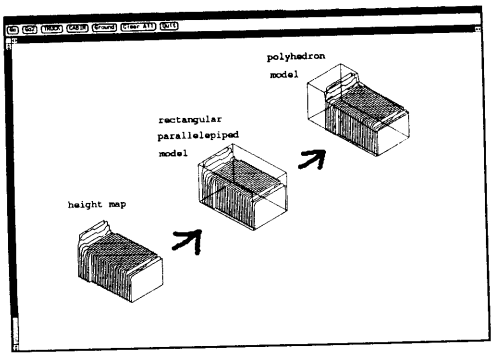
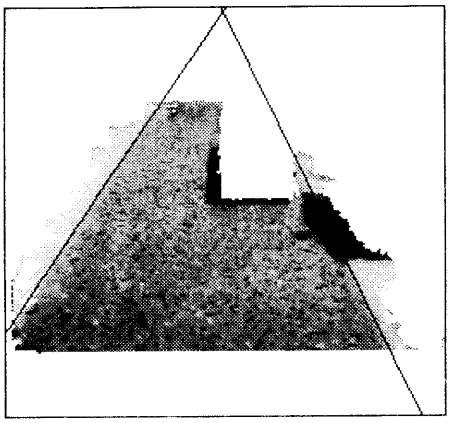
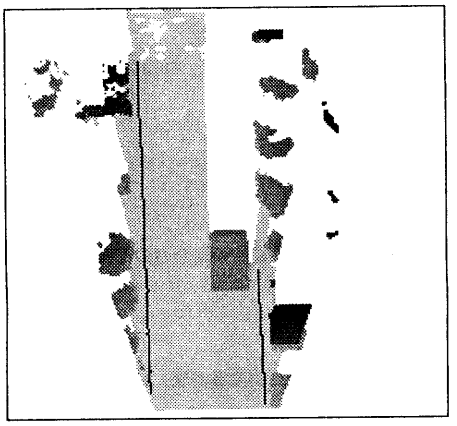


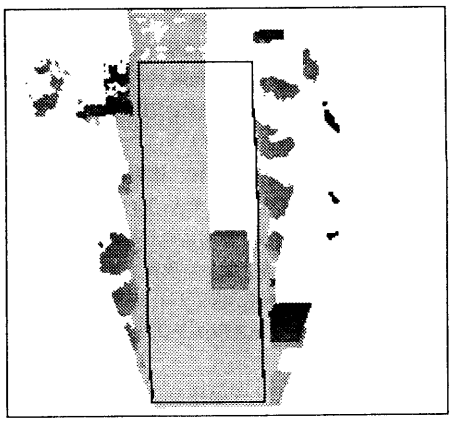
図9. トラックの場合の幾何モデルの当てはめ



(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

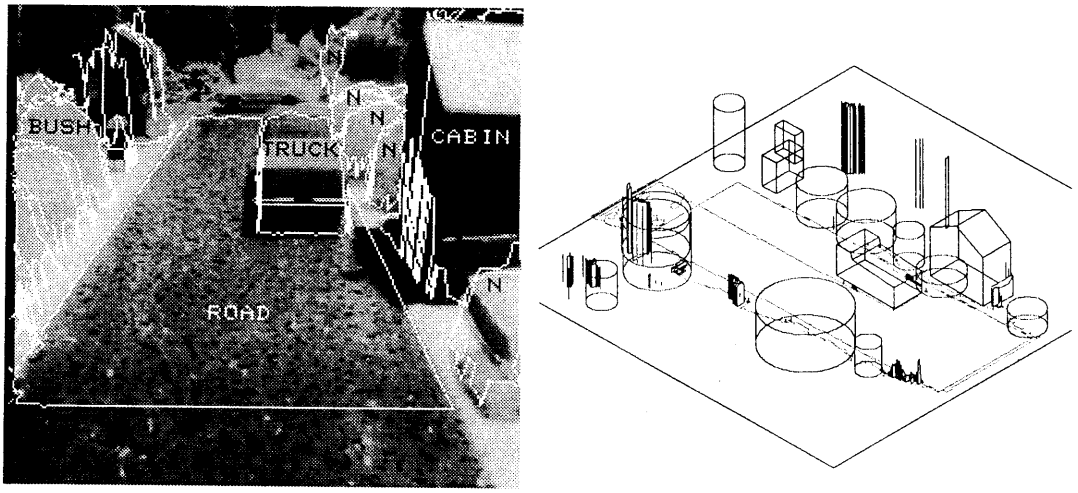


図 10. 解釈の結果と環境地図

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、トラックのモデル当てはめの過程を示す。図 10 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。

5. まとめ

センサ統合とコンピュータビジョンとの関係を明らかにし、具体例として、自律移動ロボットの場合の環境モデル構築例を示した。まとめと課題として、以下の事が挙げられる。

1) 多視点の観測から得られる冗長な情報を統計的な手法 (例えば、カルマンフィルタなど) を用いて、高精度化する手法は、十分成果を挙げており、センサ情報の下位のレベルでの融合に有効であると考えられる。それは、センサ情報の誤差分布などが事前に得られているか、あるモデルで近似できるからである。代表例としては、多視点からの位置推定などが挙げられる。

2) 環境理解などのより高度なタスクの場合には、センサ情報の不確実性が、物体の解釈などの不確実性に直接対応するとは限らない。この場合には、センサ情報の不確実性を軽減するだけでなく、センサ情報と物体の解釈に応じた環境モデル (地図) との照合が必要になる。4. で示したように、異なったセンサ情報を解釈し、それらを相補的に利用することにより、効率的かつ正しいシーンの解釈が可能になるが、相補的利用の制御構造や、環境モデルの実際の構造などに関して全てが解決した訳ではなく、より広範な研究が望まれる。

3) 多視点の観測により、センサ情報の不確実性の軽減は可能であるが、根本的には、どの地点で観測するかという「観測行動のプランニング」の問題がある。「観測のための行動 (どの地点で観測するか)」と「行動のための観測 (安全に移動できる領域はどこか)」のトレードオフをどの様に解決するかも考慮しなければならない。最近、「アクティブビジョン [20]」の研究に代表されるように、観測者の行動を制御することにより、3 次元情報の抽出を簡単化する手法が提案されている。また、移動ロボットの種々の観測形態 (観測位置と、入力画像の構成) についても研究されている [21]。これらも含めて、これから観測と行動のプランニングについて広域的に研究する必要があると考えられる。

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 Multisensor Integration, Int. J. Robot. Res., vol.7, no.6, pp.97-113, 1988.
[5] 石川、「センサヒュージョンシステム、-感覚情報の統合メカニズム-」、日本ロボット学会誌、6、3、pp.79-83, 1988.
[6] R. C. Luo and M. K. Kay: Multisensor Integration and Fusion in Intelligent Systems, IEEE Trans. Syst. Man Cybern., vol.SMC-19, no.5, pp.901-931, 1989.
[7] T. Henderson and E. Shilcrat: Logical Sensor Systems, J. Robot. Syst., vol.1, no.2, pp.169-193, 1984.
[8] T. Henderson and E. Shilcrat: The Specification of Distributed Sensing and Control, J. Robot. Syst., vol.2, no.4, pp.387-396, 1985.
[9] R. Chatila and J. Laumond: Position Referencing and consistent world modeling for the mobile robot HILARE, Proc. of IEEE Int. Conf. Robot. and Automat., pp.138-145, 1985.
[10] 片山 徹、「応用カルマンフィルタ」、朝倉書店、1983.
[11] D. J. Kriegman, E. Triendl, and T. O. Binford, "A mobile robot: Sensing, planning, and locomotion," Proc. of IEEE Int. Conf. Robot. and Automat., pp.402-408, 1987.
[12] 松田、大田、「時系列ステレオ画像の対応探索」、情処研資 CV59-2,1989.
[13] N. Ayache and O. D. Faugeras: Building, registrating, and fusing noisy visual mapps, Proc. of 1st Int. Conf. on Computer Vision, pp.73-82, 1987.
[14] A. Elfes: Using occupancy grids for mobile robot perception and navigation, Computer vol.22, no.6, pp.46-57, 1989.
[15] R. McKendal and M. Mintz: Robust fusion of location information, Proc. IEEE Int. Conf. Robot. and Automat., pp.1239-1244, 1988.
[16] A. F. Bobick and R. C. Bolles: Evolutionary approach to constructing object descriptions, Preprints of 5th Int. Symp. of Robotics Research, pp.172-179, 1989.
[17] M. Asada and Y. Shirai: Building a world model for a mobile robot using dynamic semantic constraints, Proc. of 11th Int. Joint Conf. on Artificial Intellige, pp.1629-1634, 1989.
[18] M. Asada: Builiding a 3-D world model for a mobile robot from sensory data, Proc. IEEE Int. Conf. Robot. and Automat., pp.918-923, 1988.
[19] 門野、浅田、白井、「センサ統合による道路シーンの解釈」、大阪大学知識科学研究会、第11回研究回資料、pp.1-10, 1989.
[20] D. H. Ballard: Reference frame for animate vision, Proc. of 11th Int. Joint Conf. on Artificial Intellige, pp.1635-1641, 1989.
[21] J. Y. Zheng: Dynamic Projection, Panoramic Representation, and Route Recognition, 大阪大学基礎工学部制御工学科、博士論文公聴会資料、1989年12月.