

自律走行のためのインクリメンタルな 環境モデルの構築

*佐藤 淳

* アイシン精機㈱

**築山 俊史

** 電子技術総合研究所

移動ロボットの自律的な走行のための環境モデルを、自動的に構築する手法について提案する。本手法では、まず単眼視による各画像から、そのシーンの3次元的な構造を推定し、モデルを作成する。このとき、1枚の画像からでは、モデルが一意に決定できないことがある。そのため、異なる撮像位置より得られる、同様のモデルとの対応関係を調べることにより、そのシーンのモデルを1つに決定する。移動しながらこれを繰り返すことにより、屋内全体の環境モデルの構築を行なう。

Incremental Reconstruction of the Environment Model for Autonomous Vehicles

*Jun SATO , **Toshifumi TSUKIYAMA

* AISIN SEIKI Co.,Ltd. 2-1,Asahi-machi,Kariya,448 Japan

** Electrotechnical Laboratory 1-1-4,Umezono,Tsukuba,305 Japan

In this paper a method for building a model of an indoor environment from monocular images is proposed, which will be used for an autonomous mobile robot. First, possible 3-D models for a scene of the environment are derived from the image. Then, an optimum model is decided among the candidate models by matching them to the models derived from other image of the same scene to be obtained from a different position. This procedure is repeated to get a complete model of the whole environment, while the robot is moving around.

1. はじめに

本論文では、撮像位置がかなり離れた複数枚のTV入力画像データから、撮像したシーンをモデル化する方法を提案する。提案する方法は建物内を自律走行する車のナビゲーションシステムへの応用を想定したものである。自律的な車の走行には、位置の同定や障害物回避のために、環境モデルが必要とされる。移動しながら環境をモデル化する方法として、超音波方式[1][2] やステレオ画像を使う方式[3][4][5]などが提案されている。

汎用性のためにはシーンの画像データを用いるのが望ましい。しかし詳細な解析にもとずいたシーンのモデル化には、かなりの処理時間を必要とする。自律的な走行のための環境モデルとしては、詳細な形状などの情報はあまり必要でなく、絵を書く場合のデッサン程度のスケッチができれば十分である。ここでは単眼視により得られたTV画像から、シーンの骨格となる構造を推定し、3次元モデルを作成する。一般に単眼視による画像からは、複数の3次元モデル候補が作成され、一意に決定できないという問題がある。しかし、同じシーンを異なる撮像位置から観察した場合、撮像位置は変わっても、最も確からしいモデルは似かよったものとなるはずである。そこで異なる撮像位置から得られた同様の複数の3次元モデルとの間で、最も良く一致するモデルを見つけ、1つの正しいモデルを決定する。

ところで走行中の撮像回数はできるだけ少ないほうが望ましい。しかし画像のサンプリング間隔が長くなるにつれて、対象物体の見かけの大きさや位置が大きく変わり、画像間での対応が困難になる問題が生じる。本論文では、シーンの3次元的な構造情報を使って対応関係を見つけるため、画像間での見かけの変化が大きい場合でも有効である。

2. 処理の概要

Fig. 1 に、提案するモデル化の原理を示す。カメラは時刻 t_1 に P_1 の位置からシーン内の壁と床の境界線 g （以下境界線と呼ぶ）と対象物を観測し、その位置を求める。カメラは車と共に移動し、時刻 t_2 に P_2 の位置から同じシーンを観測する。混乱を避けるため Fig. 1 では Z 軸は省略されている。提案する方式は、まず撮像位置 P_1 から得られた2次元画像から、3次元構造を推定してモデル化する。このとき単眼視により得られた1枚の画像からは、複数の3次元的解釈が可能であり、複数のシーンモデルの候補ができる。次に、撮像位置 P_2 から同様の方式で複数のシーンモデル候補を作成する。両者の対応付けにより、最も一致するモデルを選び、シーンモデルを決定する。Fig. 2 に示すように、これをロボットの走行中に繰り返し行なうことにより、環境全体のモデル化を行なう。建物内の廊下、部屋等は、通常、水平な平面（床面）と垂直な平面群（壁、柱等）でき

あがっており、またそこに存在する物体は直方体形状、または直方体に近似できるような形状である場合が多い。ここでは、対象とする空間は水平面（床）、垂直面（壁）及び直方体状の物体から成り、直方体状の物体は水平面上に置かれているものとする。シーンの構造の解析には、入力画像を処理して得られるセグメントの接合パターン[6]を手がかりに行なう。3次元情報は、まず床面を推定し、次にこれをもとに2次元画像上の座標、カメラの高さ、見込み角などから逆透視変換[6]により求める。

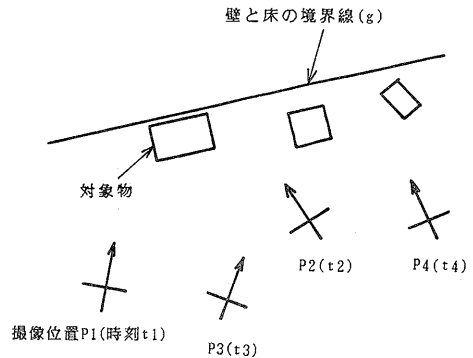


Fig. 1 モデル化の原理

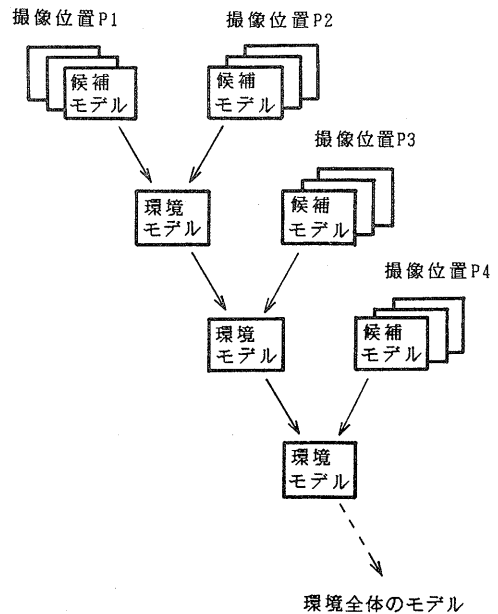


Fig. 2 環境全体のモデル化

処理の流れは次のようになる。①画像から画素の明るさの変化量をもとにエッジを抽出し、セグメント化する。②セグメントの接合パターンをもとに直方体状の物体を抽出する。③2次元画像上において壁と床の境界線を推定し、3次元シーンモデルの候補モデルを作成する。④撮像位置の異なる別の画像から得られたシーンの候補モデルとの対応により、シーンモデルを1つに確定する。⑤以上の処理を繰り返すことにより、シーンモデルを逐次更新し屋内全体の環境モデルを作成する。

3. セグメントの抽出

この節では、入力画像からエッジ点を抽出し、これから以下の4種類のセグメントを抽出する方法について述べる。

- V - Segment : 画像上で垂直な Segment
- H - Segment : 画像上で水平な Segment
- S(+)- Segment : 画像上で右上がりの Segment
- S(-)- Segment : 画像上で右下がりの Segment

ここで V, H, S(-), S(+)-Segment は物体の稜線の候補となる。

エッジの抽出は、Prewitt オペレータにより空間微分を行ない、各画素のエッジの強さと方向を求め、上記の4種類の方向に分類する。それぞれの方向ごとにエッジ点の連結性を調べ近接するエッジ点をグループ化する。各グループごとに重心を求め、重心回りのモーメントが最小となるような直線をあてはめセグメント化する。

4. 直方体の抽出

この節では、セグメントの接合パターンを利用して直方体状の物体を抽出する方法について述べる。まず線画の中から Fig.3に示すような下向き Arrow型接合点を見つかる。次にその回りに Fig.4に示すような2つのウィンドウを設け、この中に入る Vセグメントを直方体の側方の辺の候補とし、H, S(+), S(-) セグメントを直方体の上方の辺の候補とする。得られた各候補セグメントを組み合わせて、Fig.5に示すように延長し、直方体を囲んだときに、セグメントの長さが全周囲長に占める割合(P_L)が、最も大きくなるようなセグメントの組み合わせを、1つの直方体と見なす。次に、オクルージョンが起

こっていたり、線画が不完全なためにただちに直方体とは見なせないが、直方体の一部となる可能性のある垂直面を見つける。線画の中から Fig.3に示すような L型, T型, Fork型頂点を見つけ、その回りにウィンドウを1つ設けて、直方体の場合と同様の手法により垂直面を囲むようなセグメントの組み合わせを見つける。

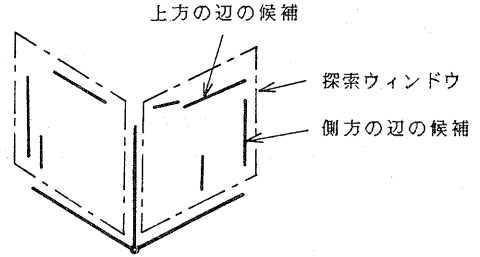
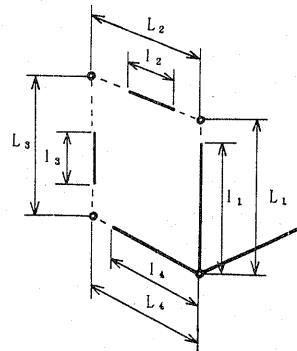


Fig.4 直方体の辺の候補の抽出



$$P_L = \frac{\sum l_i}{\sum L_i} \quad (i=1\sim 4)$$

- L_i : 直方体を囲んだときの各辺の長さ
- l_i : 直方体の辺の候補となるセグメントの長さ

Fig.5 セグメントの占める割合

5. シーンモデルの候補の作成

この節では前節で得られた直方体及び垂直面を表わすセグメントから3次元の再構成を行ないシーンモデルの候補となるモデル(候補モデル)を作成する方法について述べる。

Arrow型	L型	T型	Fork型

Fig.3 直方体、垂直面抽出に利用する接合パターン

5. 1 床と壁の境界線の推定

線画の中から床と壁の境界線を表わしていると考えられる直線を境界線候補として抽出する。すなわち境界線の存在範囲を以下の様に限定し、この条件をすべて満たす H, S(+), S(-)セグメントをすべて境界線候補とする。

(1) カメラの姿勢から、バニッシングポイントを通る水平線が求まるが、境界線はこの水平線より上に存在することは有り得ない。

(2) 境界線は2次元画像上において直方体の下を通過することは有り得ない。

(3) 直方体は不透明であるとみなせるので、2次元画像上において直方体内に存在するセグメントは境界線とは成り得ない。

5. 2 候補モデルの作成

第4節で線画を解釈して得られた直方体および垂直面のうちで、2次元画像上において、5.1節で得られた境界線より下に存在するものは、3次元空間では床面上にあると考えられる。これらの直方体については、各頂点の2次元位置データに対して逆透視変換を行ない、3次元位置データを得る。2次元画像上において境界線より下にある垂直面は3次元上では境界線より手前にあるため、直方体の一部として解釈する。境界線より上にある垂直面は壁の模様であると解釈し、直方体に含まれる垂直面は直方体の模様であると解釈する。直方体の一部として解釈された垂直面は直方体と同様に逆透視変換を行ない、3次元位置データを得る。しかし垂直面には直方体としての奥行き情報が欠落しておりこのままではモデル化できない。そこでここでは Fig.6に示すように、垂直面の奥行きとして境界線あるいは他の直方体に接するような大きさを仮定する。

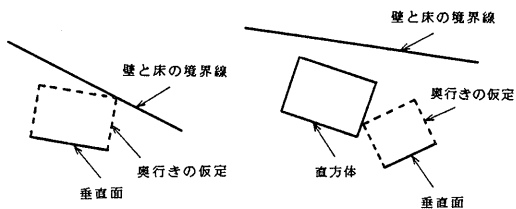
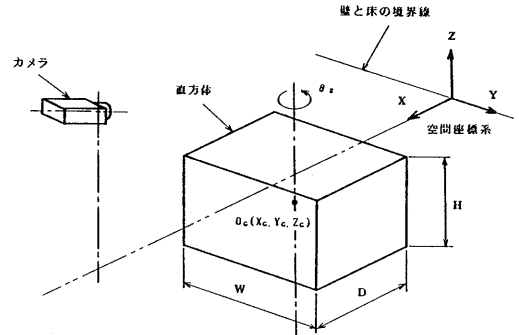


Fig.6 垂直面の奥行き仮定

次に、得られた3次元位置データから Fig.7に示す9つのパラメータを計算し、3次元ソリッドモデルである SOLVER[7] に入力しシーンの候補モデルを作成する。SO LVERは多面体、回転体等のプリミティブの作成の他、こ

れらの合成、移動、回転、投影体の作成等をおこなうことができる。ここでは直方体が水平面上に有ると仮定しているため、9つのパラメータのうち1~7までのパラメータを決定すれば良い。



1. 直方体の幅 (W)
2. " の奥行き (D)
3. " の高さ (H)
4. 直方体中心の X座標 (Xc)
5. " Y座標 (Yc)
6. " Z座標 (Zc)
7. Z軸回りの回転 (θ_z)
8. Y " (θ_y)
9. X " (θ_x)

Fig.7 直方体のパラメータ

各パラメータには、パラメータ値の他に、その値の確からしさを表わすラベルをつける。パラメータ値及びラベルは、第4節で抽出された直方体や垂直面のお互いの関係からセグメントの確からしさを推定し、これをもとに決定する。今、直方体及び垂直面が Fig.9のように抽出されたとする。S₁₁, S₁₂ が直方体R1の奥行き方向のセグメントであるとする、S₁₁, S₁₂ の長さがともに不確かな場合は、奥行きのパラメータは不確かな値しか得られないが、S₁₁, S₁₂ のどちらか一方の確かな長さが観測されれば、奥行きのパラメータは確かな値が得られる。したがってあるパラメータの値P_i及びラベルは、関係するセグメントS₁₁, S₁₂ より以下のように決定する。

1) セグメントS₁₁, S₁₂ とともに「確か」な場合、パラメータ値P_iは (S₁₁+S₁₂)/2とし、P_iには「確か」のラベルを付ける。ここで、S₁₁はセグメントS₁₁を逆透視変換した場合の3次元空間での長さである。

2) S_{11} は「確か」であるが、 S_{12} は「不確か」な場合は、 $P_i = S_{11}$ とし、「確か」のラベルを付ける。

3) S_{11} は「不確か」であるが、 S_{12} は「確か」な場合は、 $P_i = S_{12}$ とし、「確か」のラベルを付ける。

4) S_{11}, S_{12} ともに「不確か」な場合は、 P_i の値は S_{11}, S_{12} のうち大きいほうの値とし、「不確か」のラベルを付ける。

セグメントの確からしさは、そのセグメントの端点における接合パターンにより、オクルージョンを推定して判断する。またオクルードされているかどうか判定不可能な場合も、オクルードされている可能性があるとして、「不確か」のラベルを付ける。Fig. 8に、太線に注目した場合の、確からしさの判断の例を示す。Fig. 9の例では、直方体 R_1 のセグメント S_{11} は接合点 J_1 でオクルードされている可能性があり、 S_{12} は接合点 J_2 でオクルードされているため、パラメータの値としては S_{11}, S_{12} のうち大きい方を取り、「不確か」のラベルを付ける。直方体 R_2 については、 S_{21} はオクルードされている可能性があるが、 S_{22} はオクルードされていないため、パラメータの値は S_{22} とし、「確か」のラベルを付ける。

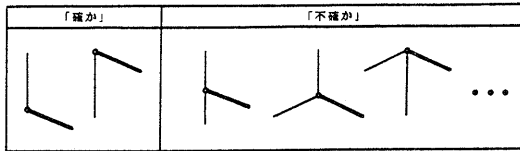


Fig. 8 セグメントの確からしさ

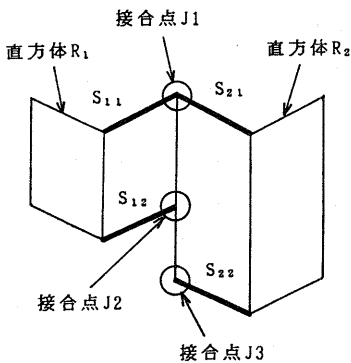


Fig. 9 パラメータの確からしさの推定

6. 候補モデルのマッチング

ある撮像位置 P_1 (Fig. 1参照)から得られた候補モデルの集合を候補モデルグループ M_1 とし、移動後の撮像位置

P_2 から得られた集合を候補モデルグループ M_2 とする。 M_1, M_2 の中の各候補モデル同士の対応をとり、最も良く一致する組み合わせを選ぶ。対応は候補モデル同士を回転及び平行移動して重ね合わせ、一致の度合いを計ることにより行なう。

今、Fig. 10に示すように、候補モデルグループ M_1 内の i 番目の候補モデル M_{1i} と候補モデルグループ M_2 内の j 番目の候補モデル M_{2j} の重ね合わせを考える。Fig. 10では混乱を避けるため、高さ方向を省略して示してある。

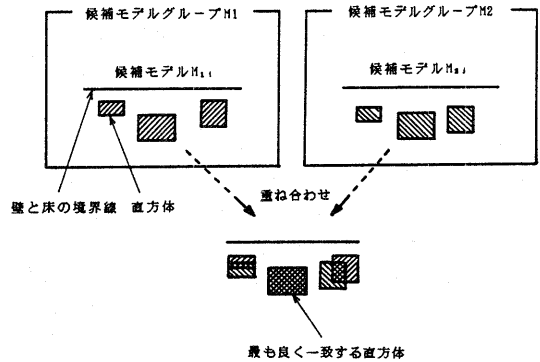


Fig. 10 候補モデルの重ね合わせ

M_1 、及び M_2 内にはそれぞれ複数の直方体が存在するが、その中から最も良く一致する直方体の組み合わせを選ぶ。直方体の一致の度合いは、お互いの直方体の持つ7つのパラメータの内、「確か」というラベルの付けられたパラメータ同士の一致の度合い(C_p)により求める。今、候補モデル M_{1i} 内の k 番目の直方体を R_{1k} 、候補モデル M_{2j} 内の m 番目の直方体を R_{2m} とする。 R_{1k}, R_{2m} の持つパラメータの内、双方共に「確か」というラベルの付いたものが n 個有り、その内の i 番目のパラメータをそれぞれ P_{1i}, P_{2i} とすると、直方体 R_{1k}, R_{2m} の一致の度合い(C_p)は式(1)により求める。

$$C_p = \frac{C_{p1} + \dots + C_{pn}}{n} \quad (1)$$

$$C_{pi} = \frac{|P_{1i} - P_{2i}|}{P_{1i} + P_{2i}} \quad (2)$$

C_{pi} : i 番目のパラメータの一致の度合い

C_P が最も小さくなるような直方体の組合わせを選び、その重心の X, Y 座標が一致するように候補モデル M_2 を平行及び回転移動して候補モデル M_1 に重ね合わせる。

次に、候補モデルの一致度 (C_M) を計算し、 C_M が最大となるような候補モデルの組合せを最適な組合せとする。 C_M は3次元の候補モデルを重ね合わせたときの積集合及び和集合の体積から以下のように計算する。

$$C_M = \frac{V_1}{V_2} \quad (3)$$

V_1 : 候補モデルを重ね合わせたときの積集合の体積

V_2 : // 和集合の体積

7. シーンモデルの作成及び更新

第6節で選ばれた最適な組合せの候補モデル同士を合成して1つのシーンモデルを作成する。合成に際しては、合成する2つの直方体の持つ各パラメータの確からしさを考慮して合成後の直方体のパラメータの値及び確からしさを決定する。合成する2つの直方体をそれぞれ R_1, R_2 とし、 R_1, R_2 の i 番目のパラメータをそれぞれ P_{1i}, P_{2i} とする。

1) P_{1i}, P_{2i} ともに「確か」の場合は、2つのパラメータの違いは計測誤差であると考え、合成後の直方体の i 番目のパラメータ P_{ci} の値は $(P_{1i}+P_{2i})/2$ とし、 P_{ci} には「確か」のラベルを付ける。

2) P_{1i} が「確か」で P_{2i} が「不確か」の場合は、 P_{ci} の値は P_{1i} とし、 P_{ci} には「確か」のラベルを付ける。

3) P_{2i} が「確か」で P_{1i} が「不確か」の場合は、 P_{ci} の値は P_{2i} とし、 P_{ci} には「確か」のラベルを付ける。

4) P_{1i}, P_{2i} ともに「不確か」の場合は、 P_{ci} の値は P_{1i}, P_{2i} のうち大きいほうの値とし、 P_{ci} は「不確か」のラベルを付ける。

7つのパラメータについてそれぞれ上述の判定を行ない、合成後の直方体のパラメータの値およびその確からしさを決定する。合成後の直方体の持つ7つのパラメータが全て「確か」となったときその直方体には「確定」のラベルを付け、そうでない場合は「不確定」のラベルを付ける。「確定」のラベルの付いた直方体は、実際に存在する直方体と考える。「不確定」のラベルの付いた直方体のパラメータの中には、「確か」のラベルが付いたパラメータも存在する。この情報を利用するため、「不確定」な直方体のデータもそのまま残しておき、次の撮像位置から得られた候補モデルとの対応時に、「確定」した直方体と共に対応をとる。撮像位置をかえながら以上の手続きを繰り返して使用することにより、実際に存在する直方体は、最初は「不確定」のラベルが付いても徐々に「確定」のラベルが付いていく。

8. 実験結果

ここで提案したアルゴリズムを検証するために室内の様々なシーンを使って実験を行った。Fig. 11にTVカメラからの入力画像の例を示す。TVカメラは高さ約75cm、見込み角約 15° であった。画像は512X480画素で各画素は256階調である。画像の入力及び処理は、TV画像入出力ボードを内蔵したSUN3/260によりおこなった。Fig. 12は入力画像から抽出された線画、Fig. 13に直方体と垂直面の抽出結果を示す。この3次元構造の解釈結果をもとに逆透視変換から3次元座標を求め候補モデルを作成した一例がFig. 14である。同様に移動後の異なる撮像位置から入力された画像Fig. 15から作成した候補モデルの一例をFig. 16に示す。そしてこれらの候補モデルから選ばれた最適な組合せについて合成しシーンモデルを作成した結果をFig. 17に示す。さらに移動しながら入力した画像Fig. 18, 19, 20からシーンモデルを逐次更新していった例をFig. 21, 22, 23に示す。移動して撮像範囲が変わるごとに、環境モデルが完成していく様子がわかる。また、ある撮像位置からはオクルージョンのために不確かな部分があったモデルも、異なる撮像位置から観察することにより、確かなモデルへと更新されている。

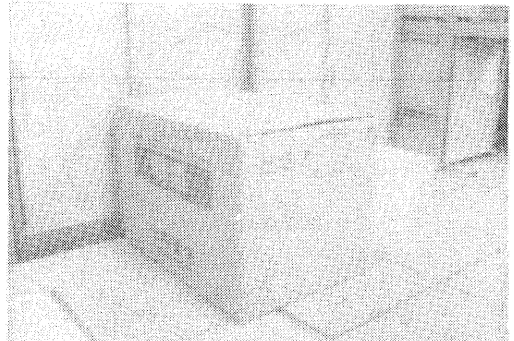


Fig. 11 TVカメラからの入力画像 1

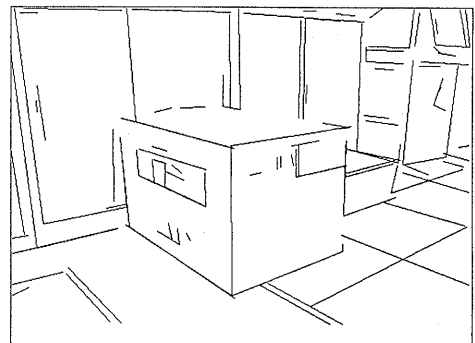


Fig. 12 抽出されたセグメント

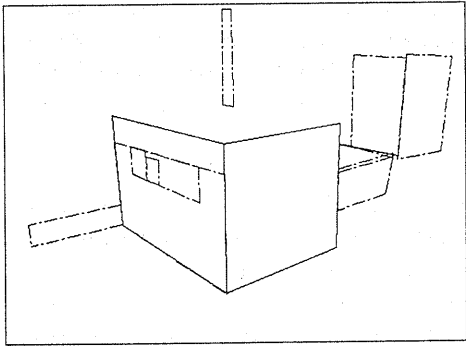


Fig. 13 抽出された直方体と垂直面

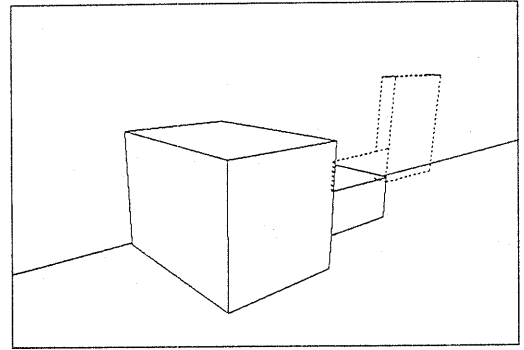


Fig. 17 シーンモデル 1

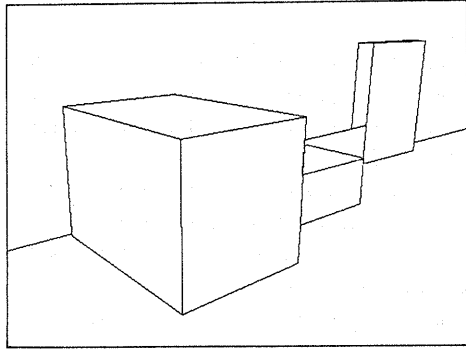


Fig. 14 候補モデルの一例

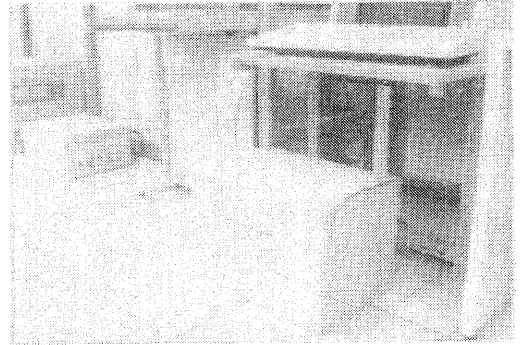


Fig. 18 TVカメラからの入力画像 3

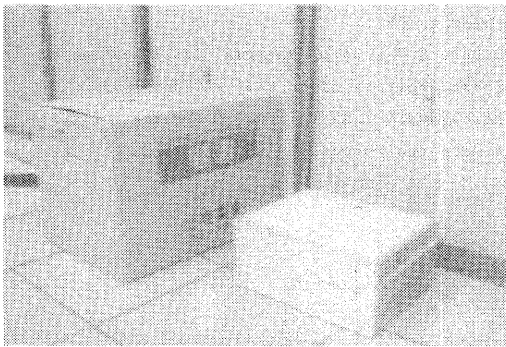


Fig. 15 TVカメラからの入力画像 2

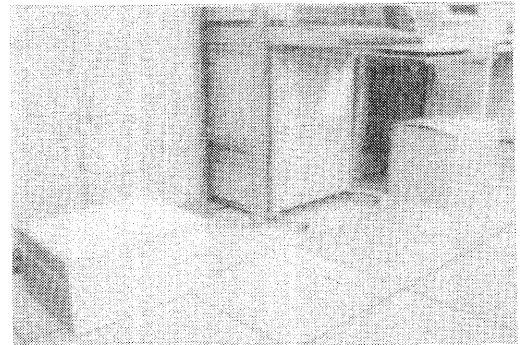


Fig. 19 TVカメラからの入力画像 4

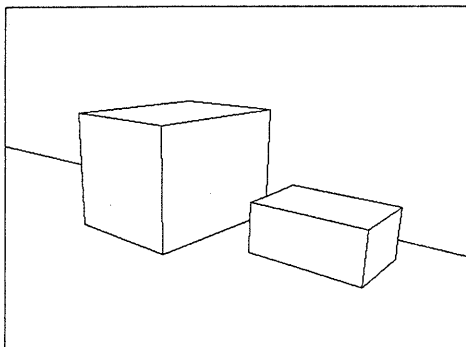


Fig. 16 候補モデルの一例

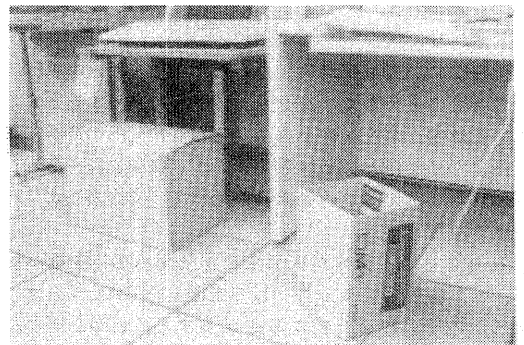


Fig. 20 TVカメラからの入力画像 5

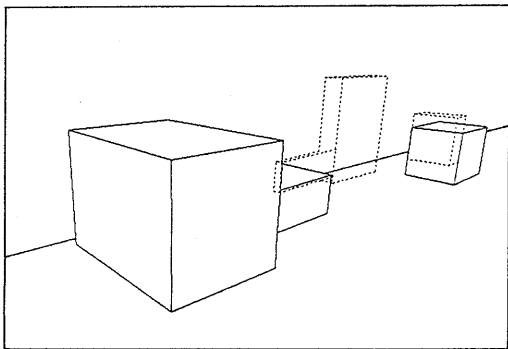


Fig. 21 シーンモデル 2

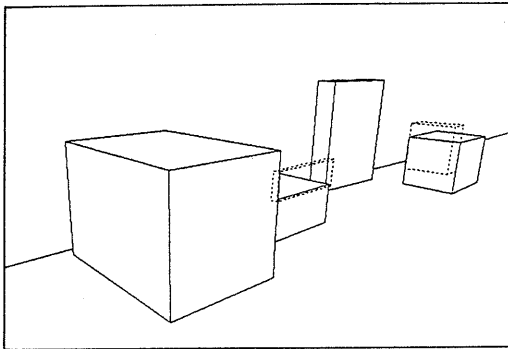


Fig. 22 シーンモデル 3

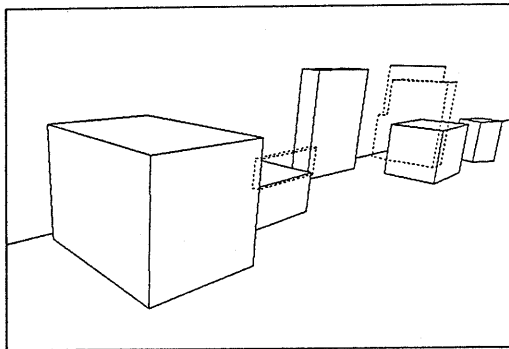


Fig. 23 シーンモデル 4

9. おわりに

屋内において車が走行しながら取り込んだ画像から、その環境のモデルをインクリメンタルに構築する方法について提案した。2次元画像からは3次元的構造が一意に決定できないという問題に対して、撮像位置を変えて得られた3次元的解釈を組み合わせて行く方式により解決した。

本論文では、物体が床面上にある場合のみを扱ったが、物体が積み重なったりして床面上を離れた場合は、何らかの手段により重なりを判別することが必要となる。これについては今後の課題である。

末筆ながら本研究を進めるにあたりご指導頂いた電総研の白井制御部長、大島室長、視覚システム研究室各位、ならびにアイシン精機㈱の関係各位に深く感謝致します。

参考文献

- [1] A. Elfes, "Sonar-based real-world mapping and Navigation", Journal of Robotics and Automation, IEEE, Vol. Ra-3, No. 3, June 1987.
- [2] J. L. Crowley, "Dynamic World Modeling for an Intelligent Mobile Robot Using a Rotating Ultrasonic Ranging Device", procs. of International Conference on Robotics and Automation, IEEE, p128-134, 1985.
- [3] E. Triendl & D. J. Kriegman, "Stereo Vision and Navigation within Buildings", procs. of International Conference on Robotics and Automation, IEEE, p1725-1730, 1987.
- [4] A. Robert de Saint Vincent, "A 3D Perception System for the Mobile Robot Hilare", procs. of International Conference on Robotics and Automation, IEEE, p1105-1111, 1986.
- [5] N. Ayache & O. D. Faugeras, "Building, Registering, and Fusing Noisy Visual Maps", procs. of First International conference on Computer Vision, IEEE, p73-82, 1987.
- [6] R. O. Duda & P. E. Hart, "Pattern Classification and Scene Analysis", John Wiley & Sons, 1973.
- [7] K. Koshikawa, & Y. Shirai, "A 3-D modeler for Vision Research", procs. of '85 ICAR, 1985.