

エッジ対応に基づく3D物体認識

玉野 隆一 佐藤 嘉伸 田村 進一

大阪大学基礎工学部

一般に、物体の3次元空間中の位置、姿勢は6自由度あるが、物体のとりうる安定姿勢を考慮に入れ回転軸を鉛直方向に限ると、平行移動を含め4つのパラメータで記述できる。この環境において、単眼の中心投影画像からそれに投影された3D物体の認識を行う。3D物体認識の問題は従来、探索の問題として扱われることが多かったが、一般化Hough変換はノイズ、隠蔽に強いという利点を持ち、それに代わる有力な並列処理手法である。本論文では、まず画像から抽出されたラインセグメントと物体モデルの稜線との仮対応から4つの変換パラメータの推定式を導き、それらの解の存在条件、個数を代数的に解析する。次にその結果を基に、一般化Hough変換により変換パラメータを決定する。この時、Hough変換の問題点であるパラメータ空間の記憶領域はとらず、各仮対応から得られたパラメータをセルとして蓄える。そしてパラメータ空間内でのピークの探索は、距離の近いセルを統合し、その大きなクラスターを求めることに置き換える。さらに得られたパラメータに対し、トップダウン処理を施し、精緻化を行う。

Three dimensional object recognition based on edge matching and generalized Hough transform

Ryuichi TAMANO Yoshinobu SATO Shinichi TAMURA

Department of Information and Computer Sciences
Faculty of Engineering Science
Osaka University
Toyonaka, Osaka 560, JAPAN

First of all, we assume the 3-D object has three translational and one vertical rotational parameters. We derive estimation formulas for four translational parameters from a tentative correspondence between line segments extracted from the image and edge lines of the object model. Then, we analyze algebraically the existence conditions and the number of the solutions. Next, based on the obtained results, we decide the translational parameters using the generalized Hough transform. In this time, we store the parameter obtained for each tentative correspondence as a cell. Further, the obtained parameters are refined by a top-down processing including reprojection to the image plane and hill climbing method.

Finally, we present an experimental result, which shows such accuracy of the proposed method that the errors are 15 cm for 10 m and 2 degrees.

1. まえがき

画像からそれに投影された既知物体の三次元空間中の位置、姿勢を決定することは、コンピュータビジョンの中心課題の1つであり、知能ロボットなどへの応用に際しても重要な技術である。一般に、物体の三次元空間中での位置、姿勢は、6つの自由度（平行移動3、回転3）があるが、物体のとりうる安定姿勢を考慮に入れ、回転軸を鉛直方向に限ると、平行移動と共に4つのパラメータで記述できる。この拘束は、現実の世界での物体の位置、姿勢を考えてもそれほど厳しいものではない。本研究で扱う問題は、単眼の中心投影画像と三次元物体モデルを照合することにより、これら4つの変換パラメータを決定することである。

画像と三次元物体モデルを照合する場合、まずボトムアップ処理により何らかの画像特徴を抽出し、その画像特徴が物体モデルのある部分に対応すると仮定して変換パラメータを推定する必要がある。仮説検定と呼ばれる方法^{[1]-[3]}では、こうして得られたパラメータをトップダウン処理により検証し、もし誤っていれば別の対応を仮定する。よってバックトラックを少なくするためには十分信頼性のある対応をどうやって見出すかが問題となる。それを解決するには、画像特徴として偶然では起きそうにない群化した特徴を選ぶ必要がある。しかし、そのような画像特徴とはどんなものなのかという問題があり、さらにノイズ、隠蔽の多い場合、その特徴が安定に抽出できるのかという問題が生じる。一方、一般化 Hough変換と呼ばれる方法^{[4]-[7]}では、抽出されたすべての画像特徴と物体のあらゆる部分との対応を仮定し、それぞれの場合の変換パラメータの仮説（以下、局所仮説と呼ぶ）を生成する。それらをすべてパラメータ空間に投影し、最も多い得票数のパラメータ値を解とする。すなわち1つ1つの局所仮説の信頼性は非常に低いが、それを蓄積し多くの証言が得られれば、その仮説は大域的につじつまの合う可能性は非常に大きくなる（以下、その仮説を大域仮説と呼ぶ）。ここで問題点は、計算時間とパラメータ空間の占める記憶容量の大きさである。しかし、画像特徴は比較的単純なものでよいので、ノイズ、隠蔽に強く、計算時間の問題も処理の並列化により解決の可能性はある。

本研究で用いる方法は一般化 Hough変換である。この考え方はBallard^[4]^[5]によって明確化された。Silberbergら^[6]とStockmanら^[7]は、物体の位置、姿勢の自由度が3（平行移動2、回転1）の環境において一般化 Hough変換を用いてこれら3つのパラメータの決定を試み、Silberbergらはさらにトップダウン処理によるパラメータの精緻化を行なっている。本研究では、平行移動の自由度をさらに1つ増やした環境を設定している。まず2.では、座標系、前提条件など問題設定を明確にする。画像特

徴としてはラインセグメント（以下、画像エッジと呼ぶ）を用い、3.において、物体モデルの稜線（以下、モデルエッジと呼ぶ）との対応を仮定したときの4つのパラメータの推定式を導出し、解の存在条件、個数などを代数的に解析する。4.では、3.の結果を基にして一般化 Hough変換により大域仮説を得る手法について述べる。ここでは、一般化 Hough変換の問題点であるパラメータ空間の記憶容量の問題を回避する手法を示す。5.では、トップダウン処理によって大域仮説における各パラメータの精緻化を行なう手法を述べ、6.では、実験結果を示す。

2. 座標系と前提条件

物体座標系 (x_m, y_m, z_m) 、ワールド座標系 (x_w, y_w, z_w) 、視点座標系 (x_v, y_v, z_v) 、画像平面座標系 (U, V) をそれぞれ図1のように設定する。物体モデルは物体座標系において頂点、直線の稜線情報からなるワイアフレームモデルで表現される。ただし、既知物体の数は1つで、画像中にその物体はただ1つ投影されているとする。物体座標系からワールド座標系への変換 T_{mw} は x_w-z_w 平面を地面と考え、平行移動 (T_x, T_y, T_z) と y 軸まわりの回転 θ の4個のパラメータで記述する。視点座標系は左手系に設定し、レンズの光軸を z_v 軸とする。ワールド座標系から視点座標系への変換 T_{vw} の回転成分は x 軸まわりの回転 α に限り、平行移動は視点座標系の原点が y_v 軸上になるように設定し、その距離を h とする。すなわち、 α は俯角、 h は視点（レンズの中心）の地面からの高さに相当し、それぞれ既知と仮定する（ただし、 h は T_y に含まれるので実験では $h=0$ とする）。透視変換 P におけるレンズの焦点距離 f およびスケールパラメータ s も既知と仮定する。画像上の点の同次座標を $i=(u, v, t)$ 、物体座標系の点の同次座標を $m=(x_m, y_m, z_m, 1)$ とし、 T_{mw} 、 T_{vw} 、 P をそれぞれ同次変換行列として表わせば^[8]、各座標系の関係は(1)式で表わされる。

$$i = m T_{mw} T_{vw} P \quad (1)$$

ただし $(U, V) = \left(\frac{u}{t}, \frac{v}{t} \right)$

$$S = \frac{s}{f} \quad T_{mw} = \begin{bmatrix} \cos \theta & 0 & \sin \theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta & 0 & \cos \theta & 0 \\ T_x & T_y & T_z & 1 \end{bmatrix}$$

$$T_{vw} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha & 0 \\ 0 & \sin \alpha & \cos \alpha & 0 \\ 0 & -h \cos \alpha & h \sin \alpha & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & S \\ 0 & 0 & 0 \end{bmatrix}$$

問題はこのような前提条件の下で、単眼の中心投影画像と物体モデルとの照合をとることにより、 T_{mw} の4つの変換パラメータ、 y 軸まわりの回転 θ と平行移動 (T_x, T_y, T_z) を決定することである。

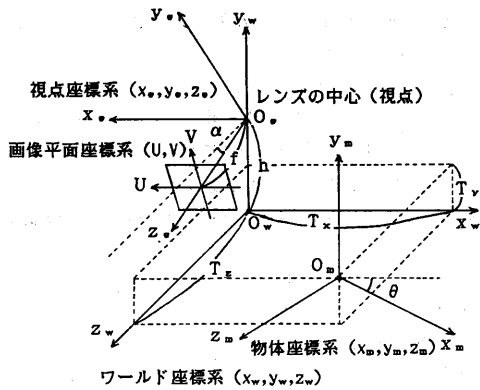


図1 座標系の設定

3. 画像エッジとモデルエッジの対応による変換パラメータの推定

(1)式において未知パラメータの数は4個である。物体上の2点とそれらの画像平面上への投影点を対応させることにより、4個の方程式が導出される。よって、これらの連立方程式を解くことにより一般に解は有限個に限定される。以下でその解法を示し、解の存在条件、個数などを代数的に解析する。

物体座標系で (x_1, y_1, z_1) - (x_2, y_2, z_2) のモデルエッジと画像平面座標系で (U_1, V_1) - (U_2, V_2) の画像エッジとの対応を仮定すると、(1)式より次の4個の式が成立する。

$$U_1 = \frac{-x_1 \cos \theta + z_1 \sin \theta - T_x}{S(x_1 \sin \theta \cos \alpha - y_1 \sin \alpha + z_1 \cos \theta \cos \alpha - T_y \sin \alpha + T_z \cos \alpha + h \sin \alpha)} \quad (2)$$

$$V_1 = \frac{x_1 \sin \theta \sin \alpha + y_1 \cos \alpha + z_1 \cos \theta \sin \alpha + T_y \cos \alpha + T_z \sin \alpha - h \cos \alpha}{S(x_1 \sin \theta \cos \alpha - y_1 \sin \alpha + z_1 \cos \theta \cos \alpha - T_y \sin \alpha + T_z \cos \alpha + h \sin \alpha)} \quad (3)$$

$$U_2 = \frac{-x_2 \cos \theta + z_2 \sin \theta - T_x}{S(x_2 \sin \theta \cos \alpha - y_2 \sin \alpha + z_2 \cos \theta \cos \alpha - T_y \sin \alpha + T_z \cos \alpha + h \sin \alpha)} \quad (4)$$

$$V_2 = \frac{x_2 \sin \theta \sin \alpha + y_2 \cos \alpha + z_2 \cos \theta \sin \alpha + T_y \cos \alpha + T_z \sin \alpha - h \cos \alpha}{S(x_2 \sin \theta \cos \alpha - y_2 \sin \alpha + z_2 \cos \theta \cos \alpha - T_y \sin \alpha + T_z \cos \alpha + h \sin \alpha)} \quad (5)$$

(2)~(5)式を展開整理すると(2)'~(5)'式のようになり、 $\sin \theta, \cos \theta, T_x, T_y, T_z$ を未知数とした線型連立方程式が導ける。

$$(SU_1 x_1 \cos \alpha - z_1) \sin \theta + (SU_1 z_1 \cos \alpha + x_1) \cos \theta + T_x - SU_1 \sin \alpha T_y + SU_1 \cos \alpha T_z = SU_1 \sin \alpha (y_1 - h) \quad (2)'$$

$$K_1 x_1 \sin \theta + K_1 z_1 \cos \theta - L_1 T_y + K_1 T_z = L_1 (y_1 - h) \quad (3)'$$

$$(SU_2 x_2 \cos \alpha - z_2) \sin \theta + (SU_2 z_2 \cos \alpha + x_2) \cos \theta + T_x - SU_2 \sin \alpha T_y + SU_2 \cos \alpha T_z = SU_2 \sin \alpha (y_2 - h) \quad (4)'$$

$$K_2 x_2 \sin \theta + K_2 z_2 \cos \theta - L_2 T_y + K_2 T_z = L_2 (y_2 - h) \quad (5)'$$

ただし、 $K_1 = SV_1 \cos \alpha - \sin \alpha$ $L_1 = SV_1 \sin \alpha + \cos \alpha$
 $K_2 = SV_2 \cos \alpha - \sin \alpha$ $L_2 = SV_2 \sin \alpha + \cos \alpha$

さらに、(2)'~(5)'式を行列を用いて書けば

$$Ax = b \quad (6)$$

となる。

ただし、

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix} \quad x = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

とおき、A、x、bの各要素は以下のようになる。

$$t_1 = \sin \theta \quad t_2 = \cos \theta \quad t_3 = T_x \quad t_4 = T_y \quad t_5 = T_z$$

$$\begin{matrix} a_{11} = SU_1 x_1 \cos \alpha - z_1 & a_{12} = SU_1 z_1 \cos \alpha + x_1 & a_{13} = 1 \\ a_{21} = K_1 x_1 & a_{22} = K_1 z_1 & a_{23} = 0 \\ a_{31} = SU_2 x_2 \cos \alpha - z_2 & a_{32} = SU_2 z_2 \cos \alpha + x_2 & a_{33} = 1 \\ a_{41} = K_2 x_2 & a_{42} = K_2 z_2 & a_{43} = 0 \\ a_{14} = -SU_1 \sin \alpha & a_{15} = SU_1 \cos \alpha & b_1 = S \sin \alpha U_1 (y_1 - h) \\ a_{24} = -L_1 & a_{25} = L_1 & b_2 = L_1 (y_1 - h) \\ a_{34} = -SU_2 \sin \alpha & a_{35} = SU_2 \cos \alpha & b_3 = S \sin \alpha U_2 (y_2 - h) \\ a_{44} = -L_2 & a_{45} = L_2 & b_4 = L_2 (y_2 - h) \end{matrix}$$

また、 t_1, t_2, θ の間には次式の関係が成立する。

$$t_1^2 + t_2^2 = 1 \quad (7) \quad \theta = \tan^{-1} \left(\frac{t_1}{t_2} \right) \quad (8)$$

ここで、

$$A_1 = \begin{bmatrix} a_{11} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{43} & a_{44} & a_{45} \end{bmatrix} \quad A_2 = \begin{bmatrix} a_{12} & a_{13} & a_{14} & a_{15} \\ a_{22} & a_{23} & a_{24} & a_{25} \\ a_{32} & a_{33} & a_{34} & a_{35} \\ a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}$$

$$B = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & b_1 \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & b_2 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & b_3 \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & b_4 \end{bmatrix}$$

とおき、以下に示すそれぞれの場合について、解の存在条件、個数とそのときのモデルエッジの幾何学的配置との対応関係を明確にする。

1) $|A_1| \neq 0$ または $|A_2| \neq 0$ の場合、(6)式は次のように変形することができる。

$$\begin{bmatrix} t_1 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} = A_1^{-1} \begin{bmatrix} b_1 - a_{12} t_2 \\ b_2 - a_{22} t_2 \\ b_3 - a_{32} t_2 \\ b_4 - a_{42} t_2 \end{bmatrix} \quad \text{または} \quad \begin{bmatrix} t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} = A_2^{-1} \begin{bmatrix} b_1 - a_{11} t_1 \\ b_2 - a_{21} t_1 \\ b_3 - a_{31} t_1 \\ b_4 - a_{41} t_1 \end{bmatrix} \quad (9)$$

(9)式より、 $t_1 = c_1 t_2 + c_2$ (または $t_2 = c_1 t_1 + c_2$) と表わせる。これを(7)式に代入することにより t_2 (または t_1) についての2次方程式が得られる。

この2次方程式の判別式をDとすると

D > 0 の時 解は2とおり存在する
D = 0 の時 解は一意に定まる
D < 0 の時 解は存在しない

解が存在するとき、そのときの t_2 (または t_1) の値を(9)式に代入することにより t_1 (または t_2)、 t_3, t_4, t_5 の値が求まる。解が存在しないときは、どのように物体を動かしてもモデルエッジ (x_1, y_1, z_1) - (x_2, y_2, z_2) は画像エッジ (U_1, V_1) - (U_2, V_2) に対応しないことを示している。

2) $|A_1|=0$ かつ $|A_2|=0$ の場合

行列式 $|A_1|, |A_2|$ の値を計算すると次式のようになる。

$$|A_1|=S\{(z_1-z_2)(V_1-V_2)+(x_1-x_2)(K_2U_1-K_1U_2)\} \quad (10)$$

$$|A_2|=S\{(x_1-x_2)(V_1-V_2)-(z_1-z_2)(K_2U_1-K_1U_2)\} \quad (11)$$

$|A_1|=0$ かつ $|A_2|=0$ となるための必要十分条件を求めると

$$x_1=x_2 \quad \text{かつ} \quad z_1=z_2 \quad (12)$$

または

$$V_1=V_2 \quad \text{かつ} \quad K_2U_1=K_1U_2 \quad (13)$$

となる。

(12), (13)式が成り立つ場合について、それぞれ 2.1) および 2.2) で述べる。

2.1) $x_1=x_2$ かつ $z_1=z_2$ の場合

地面に垂直なモデルエッジ (以下、垂直エッジと呼ぶ) を示しており解は存在しないかまたは無限に多く存在する。

解の存在条件を示す。連立方程式 (6) が解を持つための必要十分条件は $\text{rank} A = \text{rank} B$ である。そこで、行列 B に基本変形を施すと次のようになる (ただし、一般的な場合を考えて $y_1 \neq y_2, K_1 \neq 0, K_2 \neq 0$ とする^{注1)})。

$$\begin{bmatrix} -z_1+2z_2 & x_1-x_2 & 0 & d_1 & 0 & d_3 \\ x_1-x_2 & z_1-z_2 & 0 & d_2 & 0 & d_4 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

ただし、 $x_1=x_2, z_1=z_2$

$$d_1=S\left(\frac{U_1}{K_1}-\frac{U_2}{K_2}\right) \quad d_3=S\left\{-\frac{U_1}{K_1}(y_1-h)+\frac{U_2}{K_2}(y_2-h)\right\}$$

$$d_2=-\frac{L_1}{K_1}+\frac{L_2}{K_2} \quad d_4=\frac{L_1}{K_1}(y_1-h)-\frac{L_2}{K_2}(y_2-h)$$

2.1.1) $\text{rank} A=2$ の場合

$d_1=d_2=0$ より $U_1=U_2$ かつ $V_1=V_2$ が成立する。

これは、垂直エッジでその画像平面への投影像が点になる場合を示しており、画像特徴 (長さをもった画像エッジ) としてあらわれないので考慮しない。

2.1.2) $\text{rank} A=3$ の場合 ($d_1 \neq 0$ または $d_2 \neq 0$)

$\text{rank} B=3$ となるためには

$$\begin{vmatrix} d_1 & d_3 \\ d_2 & d_4 \end{vmatrix} = d_1d_4 - d_2d_3 = 0$$

が成り立つことが必要である。計算整理すると次式となる。

$$U_1 = \frac{L_1}{L_2} U_2 = \frac{SV_1 \sin \alpha + \cos \alpha}{SV_2 \sin \alpha + \cos \alpha} U_2 \quad (14)$$

この場合、(14)式を満たす画像エッジがモデルの垂直エッジと対応が可能であり (ただし、解は無限に多く存在する)、(14)式を満たさない画像エッジは物体をどのように動かしても垂直エッジに対応しない。

また、(14)式において、 α の値が小さいとき $U_1 \neq U_2$ となる。 ($\alpha=0^\circ$ のとき、 $U_1=U_2$)

$$\text{例} \quad \alpha=5^\circ \quad S = \frac{s}{f} = \frac{24}{256 \times 50}$$

$$S \sin \alpha \approx 0.000163 \quad \cos \alpha \approx 0.996$$

2.2) $V_1=V_2$ かつ $K_2U_1=K_1U_2$ の場合

$K_1=K_2$ ($\because V_1=V_2$) より $K_1(U_1-U_2)=0$ が成立する。

よって

$$V_1=V_2 \quad \text{かつ} \quad U_1=U_2 \quad (15)$$

または

$$V_1=V_2 \quad \text{かつ} \quad K_1=K_2=0 \quad (16)$$

となる。

(15), (16)式が成り立つ場合について、それぞれ 2.2.1) および 2.2.2) で述べる。

2.2.1) $V_1=V_2$ かつ $U_1=U_2$ の場合

モデルエッジの画像平面への投影像が点になる場合を示しており、画像特徴としてあらわれないので考慮しない。

2.2.2) $V_1=V_2$ かつ $K_1=K_2=0$ の場合

$$K_1=K_2=0 \quad \text{より} \quad V_1=V_2 = \frac{\tan \alpha}{S} \quad (17)$$

であるので(3)', (5)'式より $y_1=y_2, T_y = -(y_1-h) = -(y_2-h)$ が成立する。

すなわち、物体上の点 $(x_1, y_1, z_1), (x_2, y_2, z_2)$ が共にワールド座標系において視点と同じ高さにあることを示している。また、 $(U_1, V_1) - (U_2, V_2)$ は地面の無限遠線上になければならない。この場合、解は無限に多く存在する。また、無限遠線上にない画像エッジとの対応では解は存在しない^{注2)}。

4. 一般化 Hough変換による大域仮説の生成

通常の一般化 Hough変換では、各モデルエッジと全画像エッジの対応を試みて局所仮説を得、これらをパラメータ空間に投票してパラメータ空間の極大値を大域仮説とする。しかし、この場合パラメータ空間 (4次元) の占めるメモリ領域の大きさが問題点となる。本研究では、マッチングの単位としてラインセグメントを用いているため画像中の特徴の数はそれほど多くない。それに加え、安定姿勢の条件により幾何学的拘束が強く、モデルエッジと画像エッジの対応が可能な場合がある程度限られている。そこで、実際にパラメータ空間の領域をメモリ上に確保することなしに、モデルエッジと画像エッジの対応による各局所仮説を局所仮説セル (以下、セルと呼ぶ) として蓄え、セ

注1) $y_1 \neq y_2$ より $K_1=0$ かつ $K_2=0$ となることはない。

$K_1=0$ または $K_2=0$ のときも同様の方法により(14)式と同じ結果を得る。

注2) $V_1, V_2 \rightarrow \frac{\tan \alpha}{S}, T_x \rightarrow \infty$ の場合 (物体が非常に遠く

にあるとき) も考えられるが、 $T_x \ll \infty$ と仮定する。

ル間の距離を基準にセルを逐次併合することによって大域仮説を得る手法を用いる。この手法は、マッチング特徴がより高次なものとなり生成される局所仮説数がより少なくなった場合にさらに有効なものとなる。

4.1 局所仮説の生成

3. の解析によると、条件によりいくつかの場合があるが、解が有限個（高々2個）得られるのは1)の場合のみであり、大部分のエッジもこの場合に属する。また、2.1)の場合（垂直エッジ）、解は無限に多く存在するが、これに属するエッジも現実世界において数多く存在する。この他は非常に特殊な例である。よって、本研究ではこの2つの場合のみを組み込んでいる。まず、1)の場合に属するエッジの対応を基に、一般化 Hough変換を用いて局所仮説セル群を生成する。各セルには次のような情報が保存される。

Cell: $(N_i \ n_j \ T_x \ T_y \ T_z \ \theta)$

N_i : 対応するモデルエッジの番号

n_j : 対応する画像エッジの番号

T_x, T_y, T_z, θ : N_i と n_j の対応によって求めたパラメータ

セル情報には、変換パラメータと共に投票に貢献したモデルエッジと画像エッジの対応情報も含まれる。なお、垂直エッジ情報の利用については4.3で述べる。

4.2 局所仮説の統合による大域仮説の生成

4.1 で生成されたセルに対して、パラメータ空間でのセル間の距離を基準とした次のような条件を満たすセルを逐次併合することによって大域仮説を生成する。

Cell₁: $(N_{i1} \ n_{j1} \ T_{x1} \ T_{y1} \ T_{z1} \ \theta_1)$

Cell₂: $(N_{i2} \ n_{j2} \ T_{x2} \ T_{y2} \ T_{z2} \ \theta_2)$

$|T_{x1} - T_{x2}| < dT_x, |T_{y1} - T_{y2}| < dT_y, |T_{z1} - T_{z2}| < dT_z$

$|\theta_1 - \theta_2| < d\theta, N_{i1} \neq N_{i2}$

ただし、 $dT_x, dT_y, dT_z, d\theta$ はしきい値

$N_{i1} \neq N_{i2}$ の条件は同じモデルエッジ番号をもったセル間では併合を試みないことを示している。ただし、あるモデル番号 N_{i1} をもつセルに、それと異なるモデル番号 N_{i2} を持つセルが2つ以上併合されることを許すので、1つのモデルエッジに複数の画像エッジがマッチングされることがある。この修正は4.4において述べる。

上の条件は、パラメータ空間でその大きなクラスターを求めることに相当する。各大域仮説は併合されたセルのリストとして表わされ、そのパラメータの値は各セルのパラメータの値の平均値を用いる。

4.3 垂直エッジの併合

垂直エッジの対応では解は無限に多く存在する。しかし、解が存在するための必要十分条件は(14)式で与えられるので、この条件を満足する画像エッジを垂直エッジの対応候補とする。これらの対応による解は不定となるが、 θ の値が定まれば連立一次方程式(6)を解くことにより、 T_x, T_y, T_z は次式により一意に定まる。

$$T_y = \left\{ \frac{L_1(y_1 - h)}{K_1} - \frac{L_2(y_2 - h)}{K_2} - (x_1 - x_2)\sin\theta - (z_1 - z_2)\cos\theta \right\} \cdot \left(\frac{K_1 K_2}{K_1 L_2 - K_2 L_1} \right) \quad (18)$$

$$T_z = \frac{L_1(y_1 - h + T_y)}{K_1} - x_1 \sin\theta - z_1 \cos\theta \quad (19)$$

$$T_x = -(SU_1 x_1 \cos\alpha - z_1)\sin\theta - (SU_1 z_1 \cos\alpha + x_1)\cos\theta + SU_1 \sin\alpha (y_1 - h + T_y) - SU_1 \cos\alpha T_z \quad (20)$$

そこで、4.2 で得られた各大域仮説の θ の値（各セルの θ の値の平均）を用いて、各垂直エッジに対応するマッチング候補の画像エッジについて上式を用いて他のパラメータの値を計算し、垂直エッジの対応についてのセルを生成する。これらのセルを4.2と同じ条件で大域仮説に併合する。

4.4 大域仮説の修正

画像処理によって得られる画像エッジには不正確さ（特にエッジの端点の位置）が伴い、大域仮説生成の際のセル併合条件のしきい値をある程度大きく設定する必要がある。そのために1つのモデルエッジに、誤って複数の画像エッジがマッチングされてしまうことがあり、そのときにはこれを正しいもの1つに定めることが必要である。これは次のような処理によって行なわれる。

- (1) 複数の画像エッジがマッチングされたモデルエッジ N_i に対して、物体モデルにおいて N_i と頂点 V_{i1}, V_{i2} で接続しているモデルエッジでそれにマッチングされた画像エッジが1つのものを選択する。
- (2) V_{i1}, V_{i2} に対応する画像上での推定位置を、(1)で選んだモデルエッジにマッチングしている画像エッジの端点位置（ V_{i1}, V_{i2} で複数のモデルエッジに接続している場合は、それらに対応する画像エッジの端点位置の平均）から求める。 N_i にマッチングした複数の画像エッジに対してその位置に最も近い（ここでその測度は $d_1 + d_2$ で、 d_1, d_2 はそれぞれ N_i にマッチングされた画像エッジの各端点から V_{i1}, V_{i2} の画像上での推定位置までの距離とする。）画像エッジを N_i にマッチングする画像エッジとする。
- (3) これらの処理によって1つに定まらない場合は対応のすべての組み合わせを調べてそれぞれを大域仮説とする。

最後に、これら的大域仮説をマッチングしたモデルエッジの数によって多い順に並べ換える。そして、その数の最も多い大域仮説を物体モデルと最もよく対応する画像エッジ群、すなわち認識結果とする。また、その大域仮説が複数ある場合は、それぞれに対し5. で述べる精緻化を行い評価値が最小のものを認識結果とする。

5. 変換パラメータの精緻化

4. で得られた大域仮説についてトプダウン処理により精度を向上させる手法について述べる。

各大域仮説について得られたパラメータ（各セルのパラメータの平均）で、対応が求まっているモデルエッジを画像平面上に投影する。セルに保存された各モデルエッジと画像エッジの対応をもとに、投影されたモデルエッジの両端点から対応する画像エッジまでの距離の和を計算する（図2）。投影した全モデルエッジについてそれらの総和（(21)式）を評価関数とし、パラメータ空間（4次元）を山登り法で最良な評価値を求めて探索する。

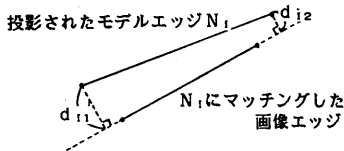


図2 投影されたモデルエッジから画像エッジまでの距離

$$\text{評価関数} : \sum_{i=1}^n (d_{i1} + d_{i2}) \quad (21)$$

ただし、 n はマッチングする画像エッジが存在するモデルエッジ数

ここで評価関数に図2に示すような距離を用いたのは画像エッジの端点の位置があまり正確でないと考えられるからである。

山登り法では大域仮説で得られたパラメータ値をパラメータ空間での初期位置とし、現在の位置に対してその近傍（ $3 \times 3 \times 3$ ）の評価値をすべて計算し（ただし、そのつどモデルエッジの投影が必要である）、評価値がより少なくなる方向へパラメータ空間を移動する。そして、その値がまわりのどの近傍の値よりも小さいとき、または繰り返し数がある回数を越えたとき探索を停止し、その時のパラメータ空間の位置を精緻化の結果とする。

6. 実験

図3のような入力画像（階段シーン）に対して、図4のような物体モデル（階段の近似モデル）を考え、その認識を試みる。各モデルエッジと6.1で述べる手法で得られた画像エッジに対して、4. で述べた手法で、局所仮説およびその統合により大域仮説を生成し、5. で述べた手法により精緻化を行なった結果を示す。

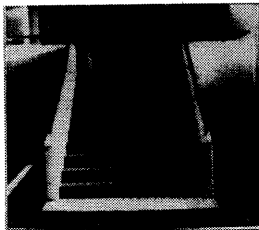


図3 入力画像

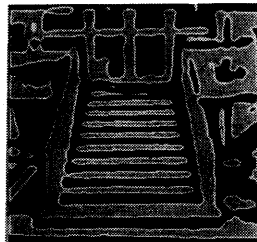


図5 ゼロ交差輪郭

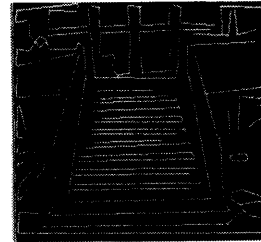


図7 直線近似（フィルタリング処理後）

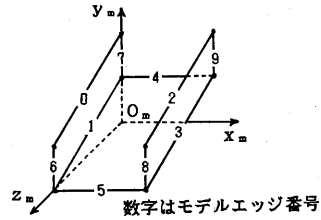


図4 物体モデル（階段の近似モデル）

6.1 画像エッジの抽出

6.1.1 エッジ点の検出

入力画像は 512×512 画素（各点8bit）で、それを 257×257 画素にスムージングおよび縮小し、さらにBurtの手法^[5]によりDOG（Difference-of-Gaussian）フィルタを適用する。DOGフィルタの出力画像（以下、DOG画像と呼ぶ）のゼロ交差点をエッジ点とし、各ゼロ交差輪郭（閉曲線、図5）をエッジ点列として格納する。

6.1.2 直線近似

得られた各エッジ点列について、split法により最大誤差 ϵ 以下で多角形近似する。これは、各点列についてその両端を結ぶ直線から最も距離の離れた点を分割点とし、その距離が ϵ 以下になるまで分割を繰り返す手法である。求めた各線分について、距離の2乗和を評価関数にして最小2乗法を適用して直線近似し、初期ラインセグメント集合を得る。

一般に、ゼロクロッシングは非常に弱い変化も検出してしまう。また、split法のみでは適切な近似直線を得ることができない。そこで、次のようなセグメント単位のフィルタリング処理により初期ラインセグメントの修正を行なう。

まず、非常に弱い変化を除去するためにDOG画像にSobel微分オペレータを適用し、各ラインセグメント上でその微分値の平均と総和があるしきい値以下のセグメントを初期集合から除去する。次に、残りのセグメントに対して以下の条件を満たすセグメントを逐次併合する。

条件：セグメント l_i について両端点を中心に $dr \times dw$ のウィンドウを考えて、このウィンドウ内に端点のどちらか一方が入っており、かつ $|\theta_{l_i} - \theta_{l_j}| < d\theta$ を満たすセグメント l_j を l_i に併合する。(図6 (a))

ただし、 $\theta_{l_i}, \theta_{l_j}$ はセグメント l_i, l_j の傾きで、短いセグメントは角度差が大きくても併合するようにしきい値 $d\theta$ の値を、セグメント l_i, l_j の長さ $leng_{l_i}, leng_{l_j}$ としきい値 d によって次のように変化させる^[10]。

$$l = \min(leng_{l_i}, leng_{l_j})$$

$$\text{if } l < d \text{ then } d\theta = 90^\circ$$

$$\text{else } d\theta = \max(d\theta_1, \tan^{-1}\left(\frac{d}{\sqrt{l^2 - d^2}}\right))$$

また、 l_i と併合するセグメント l_j がすでに別のセグメントに併合されている場合がある。その時は、併合されているセグメントの中で最も長いセグメント l_k について l_k の傾き方向に幅 dw の領域を考え、 l_i の端点のどちらか一方がこの領域内に入っている時、併合する(図6 (b))。これらの処理終了後、併合されたセグメントに対して再び最小2乗法を適用し直線近似する。

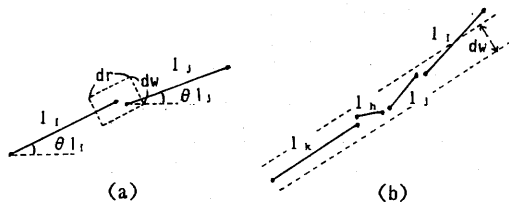


図6 ラインセグメントの併合条件

以上の処理によって得られたラインセグメント集合を実際にマッチングを行なう画像エッジとする。

図7に直線近似の結果(フィルタリング処理後)を示す。得られた画像エッジの数は131本であった。

6.2 認識結果

物体モデルエッジと6.1で得られた画像エッジに対して、4. で述べた手法を適用した。生成された局所仮説数は1252個で、2個以上の局所仮説が統合された大域仮説を表1に示す。得られた大域仮説の中で、マッチングした画

表2 修正後の大域仮説

	モデルエッジ番号									変換パラメータ					
	0	1	2	3	4	5	6	7	8	9	T_x	T_y	T_z	θ	
大域仮説番号	1	58	31	52	37	-	36	57	30	53	38	120	-160	992	177
	2	52	37	-	-	-	-	53	38	-	-	-84	-155	999	178
	3	-	58	31	-	-	-	-	53	18	-	126	-124	939	178
	4	-	-	58	31	-	-	-	57	30	-	318	-160	982	176
	5	52	53	-	-	-	-	46	-	-	-	-98	-203	596	111
	6	-	-	52	53	-	-	-	46	-	-	-295	-203	555	111
	7	-	-	-	58	-	-	105	-	-	20	-138	-141	702	345
	8	-	5	-	9	-	-	-	-	-	-	300	216	3218	175
	9	-	-	37	-	-	-	-	-	-	-	269	214	3224	176
	10	31	-	37	-	-	-	-	-	-	-	113	-224	910	179
	11	-	5	-	9	-	-	-	-	-	-	300	286	3218	179
	12	-	7	-	9	-	-	-	-	-	-	269	284	3224	176

数字は画像エッジ番号

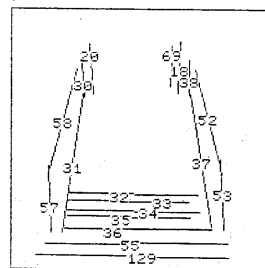
像エッジが存在するモデルエッジの数が最も多いものについて、マッチングした画像エッジを図8に示す。4.4で述べた修正後の各大域仮説およびパラメータの値を表2に示す。大域仮説の中で、モデルエッジのマッチング数が最も多いものについて精緻化過程を表3に、精緻化前後の結果を図9に示す。また、別の方向から撮影された画像に対して精緻化後の結果のみを図10に示す。図9と図10の処理結果について、パラメータ値の計算値と実測値を表4に示す。

システムは、VAX-11/780 (VMS) 上に構築されており、C言語で記述されている。なお、処理時間は画像エッジの抽出に約1分40秒、認識に約40秒である。

表1 生成された大域仮説(修正前)

併合されたセルのリスト	
1	((0 58)(1 31)(2 52)(3 37)(5 32)(5 33) (5 34)(5 35)(5 36)(5 55)(5 129)(6 57) (7 20)(7 30)(8 53)(9 18)(9 38)(9 69))
2	((0 52)(1 37)(6 53)(7 18)(7 38)(7 69))
3	((1 58)(3 52)(8 57)(9 18)(9 69))
4	((2 58)(3 31)(8 57)(9 20)(9 30))
5	((0 52)(1 53)(7 46))
6	((2 52)(3 53)(9 46))
7	((3 58)(5 105)(9 20))
8	((1 5)(1 7)(3 9))
9	((0 31)(2 37))
10	((0 5)(0 7)(2 9))

セルは(モデルエッジ番号 画像エッジ番号)として表示されている。

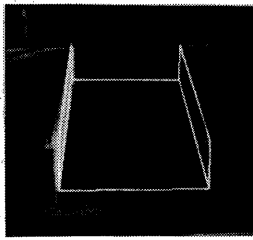


数字は画像エッジ番号

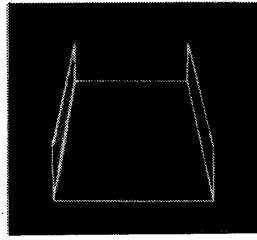
図8 マッチングした画像エッジ

表3 精緻化過程

移動回数	T_x	T_y	T_z	θ	評価値
0	120	-160	990	177	106.6
1	118	-162	990	178	99.4
2	116	-164	970	179	77.6
3	116	-166	960	180	68.1
4	116	-168	950	179	64.8



(a)



(b)

図9 精緻化の結果

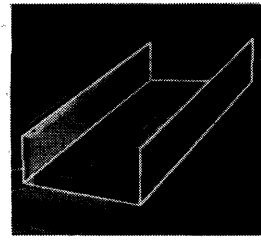


図10 精緻化の結果

6.3 考察

表1の実験結果によると、平行移動で最大15cm、回転で 2° 程度の誤差が生じた。これはカメラと物体間の距離が約10m弱で、また近似モデルを用いていることを考えると満足のいく結果である。また、背景などを含む多くのエッジの中から物体を見つけだすことができた。

局所仮説の統合には、この場合15秒程度の処理時間が必要であるがこれはセルのすべての組み合わせを調べているからであり、セル数が少なくなれば処理時間も短縮される。そこで自由度を3(T_y を固定し、視点の地面からの高さを既知とする)にすると、さらに幾何学的拘束が強くなり生成される局所仮説数がより少なくなり、本手法はさらに有効なものとなる。また、この条件も現実世界の物体を考えるとそれほど厳しい拘束ではない。

精緻化処理では、初期値が最良値に近いので山登り法の問題点(局所最大、高原、尾根)は問題にならず数回の探索により最良値に到達することができた。

7. むすび

3D物体の位置、姿勢を決定するために、安定姿勢の拘束条件の下で物体モデルエッジと画像エッジの仮対応をとることにより変換パラメータの推定式を導き、解の存在条件、個数などを代数的に解析にした。また、一般化 Hough 変換による物体の認識手法、変換パラメータの精緻化手法を示した。

本研究では、既知物体の数は1つで、その物体は画像中でただ1つ投影されていると仮定したが、今後の課題としては複数の既知物体があり、画像中に投影されている物体数が未知の場合が考えられる。この場合、大域仮説の生成までは、各局所仮説セルの情報に物体ラベルを付加することにより容易に拡張できるが、各大域仮説の検証法を考慮する必要がある。また、内部パラメータ(物体稜線の長さ、稜線間の角度)を持つ3D物体の認識、物体表現および照合法の階層化なども重要な課題であり現在検討している。

表4 変換パラメータの計算値と実測値

入力 画像	計算値				実測値			
	T_x	T_y	T_z	θ	T_x	T_y	T_z	θ
図9	116	-168	950	179	111	-170	960	180
図10	-18	-156	970	217	-20	-160	984	219

単位 距離: cm 角度: deg

参考文献

- [1] Tropsf, H., and Walter, I.: "An ATN model for 3-D recognition of solids in single images", Proc. of 8th IJCAI, pp.1094-1098(1983).
- [2] Bolles, R.C., Horaud, P., and Hannah, M.J.: "3DP0: A three-dimensional part orientation system", Proc. of 8th IJCAI, pp.1116-1120(1983).
- [3] Lowe, D.G.: "Visual recognition from spatial correspondence and perceptual organization", Proc. of 9th IJCAI, pp.953-958(1985).
- [4] Ballard, D.H.: "Generalizing the Hough transform to detect arbitrary shapes", Pattern Recognition vol.13, no.2, pp.111-122(1981).
- [5] Ballard, D.H., and Sabbah, D.: "On Shapes", Proc. of 7th IJCAI, pp.607-612(1981).
- [6] Silberberg, T.M., Harwood, D., and Davis, L.S.: "Three dimensional recognition using oriented model points", in Techniques for 3-D Machine Perception, Rosenfeld, A. (ed.), North-Holland, pp.271-320 (1986).
- [7] Stockman, G., and Esteva, J.C.: "Use of geometrical constraints and clustering to determine 3D object pose", Proc. of 7th ICPR, pp.742-744(1984).
- [8] 山口富士夫: "コンピュータディスプレイによる図形処理工学", 日刊工業新聞社, pp.135-211(1981).
- [9] Burt, P.J., and Adelson, E.H.: "The laplacian pyramid as a compact image code", IEEE Trans., vol.COM-31, no.4, pp.532-540(1983).
- [10] Medioni, G., and Nevatia, R.: "Segment-Based Stereo Matching", CVGIP, vol.31, no.1, pp.2-18(1985).