

ソリッドモデルと画像の定性的および定量的照合

小泉 正彦*

*松下電器産業(株)東京研究所

高田 文明**

**電子技術総合研究所制御部

ソリッドモデルと画像を照合する手法を紹介する。照合の手順は2段階に分かれている。まず最初に、画像に投影されている物体の候補となり得るモデルを選ぶ。そのための手がかりとして、物体を階層を持つ要素に分解したプリミティブを使う。画像から定性的特徴によってプリミティブを見つけることでモデルの候補を得る。これを定性的照合と呼ぶ。次に、選ばれたモデルの候補と画像とを照合する。局所的な特徴によって初期照合を行い、その結果を全所的な特徴によって検証し、さらに微調整を行って正しい結果を得る。これを定量的照合と呼ぶ。座標変換の回転成分の直行問題を回避するための新しい方法を紹介する。また本手法に基づいて複雑なシーン中に置かれた物体に関して、実験を行い良好な結果を得たので報告する。

Qualitative and Quantitative Matching of Solid Models and Image of 3D Objects

Masahiko KOIZUMI*

*Tokyo Lab., MATSUSHITA Electric Ind. Co., Ltd.
Higashimita 3-10-1, Tama-ku, Kawasaki-shi, 214

Fumiaki TOMITA**

**Control Division, Electrotechnical Laboratory
Umezono 1-1-4, Tsukuba-shi, 305

This paper presents a model-based vision system which recognizes 3D (three-dimensional) objects in an image by matching 2D boundary segments in an image with 3D boundary segments in a model. We introduce a new method to get a 3D coordinate transformation (scaling, translation, and rotation) from an object model to image data. A candidate transformation is initially hypothesized by matching a few segments locally, and it is tested and adjusted with the rest of segments for verification. However, the search space for the initial quantitative match is very large even when recognizing a simple object. We solve this problem by decomposing object models into common primitives and by detecting views of primitives qualitatively in the image.

1. まえがき

本報告では、物体のソリッドモデルと画像とを照合する一手法を紹介する。

照合は、物体のモデルと画像との対応点を求め、それらの間の座標変換を求める問題である。ところが、たとえ認識しようとする物体が単純なものであったとしてもシーンが複雑であると、最初に対応点を求めることは困難である。さらに実際には、多くのモデルと多くの姿勢があり、その中から適切なモデルと姿勢を選択しなければならない¹⁾²⁾。

われわれの提案する照合の手順は2段階に分かれている。まず最初に多くのモデルの中から、画像に投影されている物体との照合の候補となり得るものを選択するステップがある。ここではモデルおよび画像の定性的な特徴を用いる。このステップを定性的照合³⁾と呼ぶことにする。われわれは複雑な物体を認識するための手がかりとなるプリミティブを考えた。複雑な物体を前もってプリミティブに分解しておき、画像中のプリミティブを定性的特徴によって探索する。第3章では平行線、平行四辺形、および直方体を画像のプリミティブとして抽出する並列アルゴリズムを紹介する。

次に、選ばれたモデルの候補と画像とを照合するステップがある。ここではモデルおよび画像の定量的な特徴を用いて照合を行う。この照合を定量的照合⁴⁾と呼ぶことにし、第4章で述べる。まず最初に局所的な特徴を照合することによって座標変換の初期値を決める。その後、残りの全体の特徴に対して初期値として求めた座標変換を適用することによって正当性を検証し、さらに正確な変換を求めるための微調整を行う。

なお物体の画像への投影は、スケールだけが距離に反比例して変わる正規化透視投影⁵⁾と仮定する。

2. データ構造

2.1 ソリッドモデル

物体のモデルは B-rep (Boundary Representation; 境界表現) によって表される。図1はある物体のデータ構造を示す例である。物体Oは面Fのリストとして表され、それぞれの面Fは境界(外周境界Bと、内部

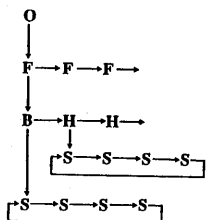


図1. 物体を表すモデルのデータ構造

に穴がある場合にはその穴との内周境界H)のリストとして表され、さらにそれぞれの境界は面を自分の右側に見るような方向を持った境界線セグメントSのリストによって表される。それぞれのセグメントは線分もしくは円弧として表される。このような表現はソリッドモデルや物体の距離データを処理することによって生成することができる⁶⁾。

2.2 画像データ

われわれはすでに3次元の B-rep の仕様にしたがって、画像を表現するための2次元の B-rep を開発した⁷⁾。これは画像の中間表現として用いられる。この表現はもとの画像を再構成するのに十分な情報を持っているばかりでなく、多くの高レベルでの後処理に有効に使うことができる。

図3(a)と3(b)に示すように、エッジ検出法によって画像は領域に分割される。それぞれの領域Rは外周境界Bといくつかの穴との内周境界Hによって表される。そして外周境界B、内周境界Hともに図2(a)および2(b)に示されるような、領域を自分の右側に見るような方向を持った境界線セグメントSのリストによって表される。モデルのセグメントが3次元の線分あるいは3次元の円弧であったのに対応して、画像のセグメントは線分もしくは楕円弧によって近似される。図3(c)は画像から得られた境界線を示している。

2.3 ρ - θ リスト

画像の線分セグメントは、図4に示すような直線の式 $\rho = x \cos \theta' + y \sin \theta'$ と線分の向きとによって ρ - θ 配列に分類される。すなわち ρ - θ 配列の要素それぞれは、同じ ρ - θ 値を持つ線分セグメントのリストである。これを「 ρ - θ リスト」と呼ぶことにする(要素がセグメントの長さの総和であれば普通の Hough 変換である)。 ρ - θ リストは、定性的照合において平行な線分セグメントや同一直線上にある線分セグメント等の、画像のプリミティブを抽出するのに使われる。また定量的照合においてモデルのセグメントに相当するはずの画像中のセグメントを効率的に発見するのに使わ

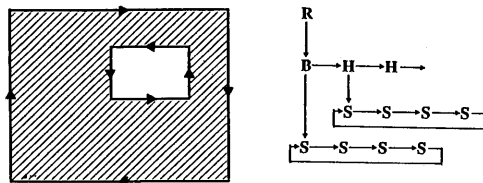


図2 (a) 領域とその境界 図2 (b) 領域を表すデータ構造

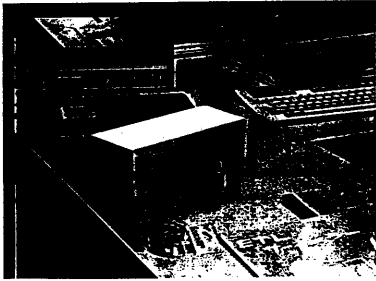


図3 (a)入力画像

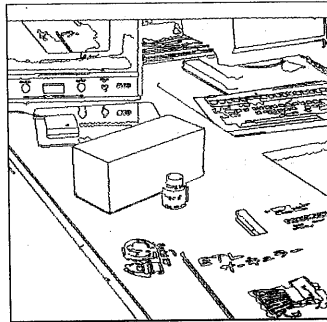


図3 (b)エッジ画像

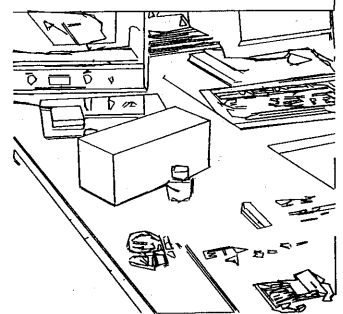


図3 (c) 2次元境界表現

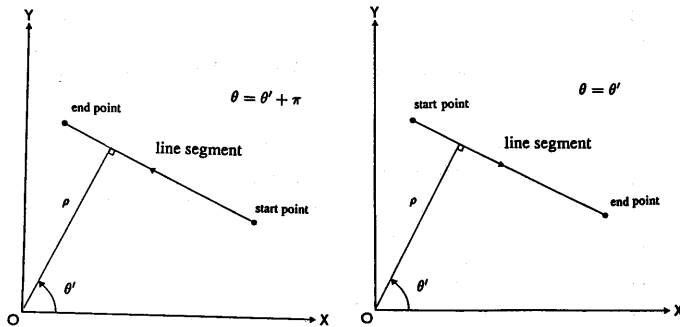


図4. 線分セグメントの ρ - θ 表現

れる。

3. 定性的照合

定性的照合は画像に投影されている物体との照合の候補となり得るモデルを選び出すステップである。そのために物体のモデルを単純なプリミティブにあらかじめ分解しておいて、画像中でのプリミティブを発見する。プリミティブとは複雑な物体を認識するための手がかりとなるものである。詳細な照合は定量的照合の段階で行われるので、物体の全ての要素がプリミティブである必要はない。

プリミティブにはいくつかのレベルがある。例として図5に示すようなテレビ受像機について考える。テレビ受像機は直方体に分解される。直方体は三つの平行四辺形に分解される。平行四辺形は二組の平行線に分解される。これら平行線、平行四辺形、および直方体はみなプリミティブである。プリミティブは、その絶対的な長さや角度を考慮しないで、その定性的特徴によって表されている。結果として、高レベルのプリミティブが少数の物体によって共有されるが、低レベルのプリミティブはいくつもの物体によって共有され

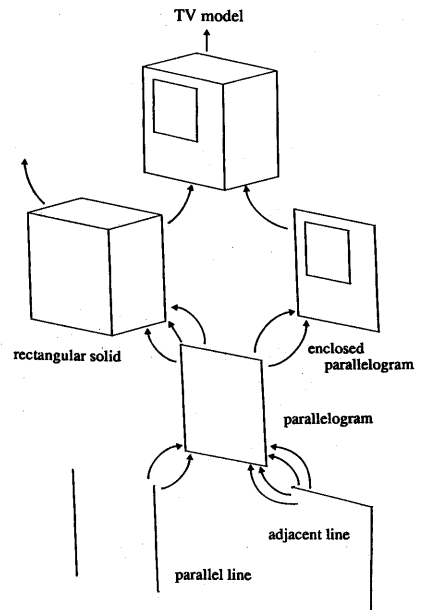


図5. テレビ受像機の階層

ることになる。低レベルになるほどプリミティブは発見するのが簡単であり、実際探索の早いうちに発見される。低レベルのプリミティブが発見されると、それより高いレベルのプリミティブの発見が行われるようになり、そして最後には物体モデルを呼び出す。その結果、定量的照合はその後の狭められた探索空間の中で行うことができる。異なったプリミティブが同じレベルのプリミティブや物体を導いたり、また一つのプリミティブが二つ以上のプリミティブや物体に関わっていたりする。

本章では、画像中の平行線、平行四辺形、および直方体のようなプリミティブを抽出する並列アルゴリズム

ムを紹介する。画像にはオクルージョンやノイズがありまた物体の表面には模様があるため、完全なプリミティブを最初から発見することは難しい。例えば図3の直方体はカメラのレンズによって一部分が隠されている。われわれの提案するアルゴリズムはそのような不完全な画像から完全なプリミティブを構成するのである。

3.1 セグメント間の関係

まず最も低いレベルのプリミティブについて述べる。例えば平行四辺形を発見するためには、セグメント間の関係を表す2種類のプリミティブが用いられる。それらは ρ - θ リストから導かれるもので、P-リストとC-リストと呼ぶ。

① P-リスト

互いに平行な線分セグメントの対のリストである。二つのセグメント S_i 、 S_j の ρ - θ の値がそれぞれ (ρ_i, θ_i) 、 (ρ_j, θ_j) であるとする。もし $\rho_i \neq \rho_j$ かつ $\theta_i = \theta_j + \pi$ であれば S_i と S_j はP-リストに加えられる。ここで $\theta_i = \theta_j$ の場合には S_i と S_j はP-リストに加えられることはない。それはセグメントには領域を自分の右側に見るように方向がつけられているので、これらのセグメントの対は凸の領域を囲まないからである。図6は図3において発見された平行線セグメントを示している。

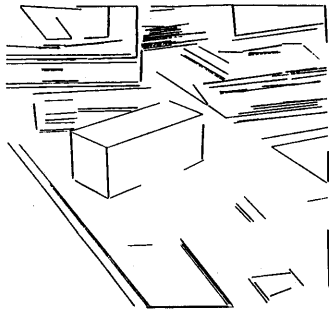


図6. 図3(c)から発見された平行線

② C-リスト

凸多角形の2辺となることのできるセグメントの対のリストである。もし一方のセグメントの延長が他方のセグメントに直接に交わるのであれば、それらのセグメントは凸の多角形を作ることはできない。図7の二つのセグメント S_i と S_j において、もし以下の条件が満足されれば S_i と S_j はC-リストに加えられる。

$$\begin{aligned} |\vec{a} \times \vec{b}| &< 0, \\ |\vec{b} \times \vec{c}| &< 0, \\ |\vec{c} \times \vec{d}| &< 0, \\ |\vec{d} \times \vec{a}| &< 0. \end{aligned}$$

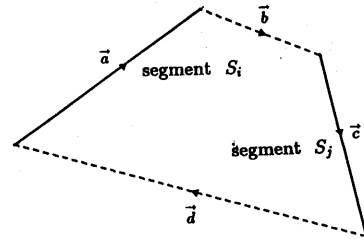


図7. 凸多角形の2辺の候補となるセグメント

3.2 平行四辺形の発見

平行四辺形はP-リストおよびC-リストを用いることで図8のセグメントA、B、C、Dとして発見される。

- ① P-リストの中から(A, C)を見つける。
- ② C-リストの中から(A, B)を見つける。
- ③ C-リストの中に(B, C)が存在することを確認する。
- ④ C-リストの中から(C, D)を見つける。
- ⑤ C-リストの中に(D, A)が存在することを確認する。
- ⑥ P-リストの中に(B, D)が存在することを確認する。

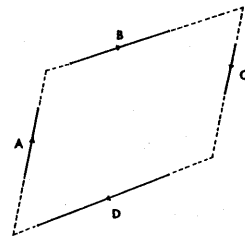


図8. 平行四辺形の4辺の候補となるセグメント

この手続きでは、図9に示すように平行四辺形の一つの辺が二つ以上のセグメントに分れている場合、同じ平行四辺形が異なったセグメントの組として発見されることになる。そのためすべての二つの平行四辺形の組A、B、C、DおよびA'、B'、C'、D'について、それらが同じ平行四辺形を表しているかどうかをテストしなければならない。もし以下の条件が満足されれば、その二つの平行四辺形は同じものであり、マージされる。

($A = A'$ or A が A' と同一直線上にある) and
 ($B = B'$ or B が B' と同一直線上にある) and
 ($C = C'$ or C が C' と同一直線上にある) and
 ($D = D'$ or D が D' と同一直線上にある)

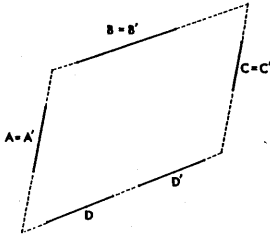


図9. 1辺が二つ以上のセグメントに分割された平行四辺形

最後に、発見された平行四辺形の確からしさがテストされる。1を平行四辺形を構成するセグメント S_i の長さ、Lを平行四辺形の周長とする。確からしさは

$$\frac{\sum l_i}{L}$$

によって定義する。もしその値が閾値を越えれば、平行四辺形と見なす。閾値を越えない場合は、セグメントが発見されていない部分を探索して平行四辺形の辺となり得るセグメントを探す。この探索を行っても閾値を越えないのであれば、それは平行四辺形とは見なされない。図10は図3から発見された平行四辺形を示す。

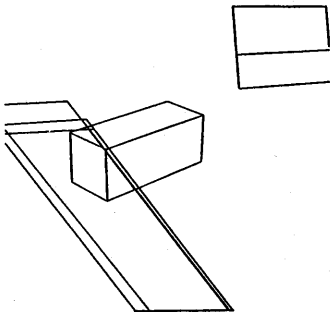


図10. 図3(c)から発見された平行四辺形

3.3 直方体の発見

もし三つ以上の平行四辺形が発見されたならば、その中のどれか三つの平行四辺形の組が直方体を構成す

るかをチェックする。図11に示すように、 P_1 、 P_2 、 P_3 を平行四辺形とする。もし以下の条件が満足されれば、その三つの平行四辺形は直方体を構成する。

P_1 と P_2 の間に共通な辺 S_{12} が存在する。

P_2 と P_3 の間に共通な辺 S_{23} が存在する。

P_3 と P_1 の間に共通な辺 S_{31} が存在する。

S_{12} 、 S_{23} 、 S_{31} の間に共通な点が存在する。

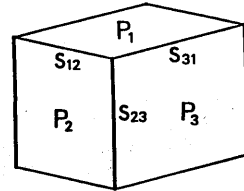


図11. 直方体を作る三つの平行四辺形

この条件だけでは平行六面体も発見されることになる。しかしその区別は定量的照合の段階で明らかになるので問題とはならない。図12は図3において発見された直方体を示している。

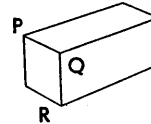


図12. 発見された直方体

3.4 実験結果

図13はテレビ受像機の濃淡画像である。その2次元B-repを図14に示す。図14からテレビ受像機の外形(=直方体)を発見した結果を図15に示す。

また図16は表面に模様のあるティッシュペーパーの箱の画像である。図17はその2次元B-rep画像、図18は発見されたティッシュペーパーの箱を示す。このように表面に模様等がある場合にも発見することができる。

4. 定量的照合

定量的照合における問題は3次元の物体のモデルを2次元の画像に一致させるような座標変換を決定することである。通常は最小二乗法⁹⁾などによってこの座標変換を行列の形で得る。しかしこの場合、回転成分

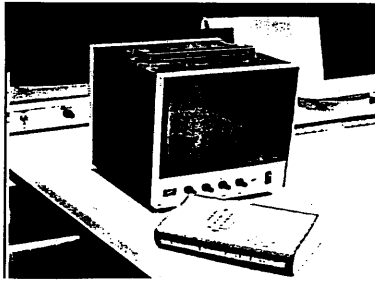


図13. 入力画像(テレビ受像機)

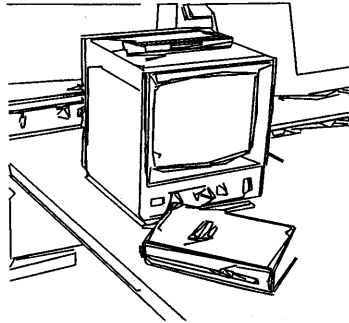


図14. 図13の2次元境界表現

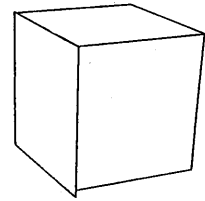


図15. 発見されたテレビ受像機の外形

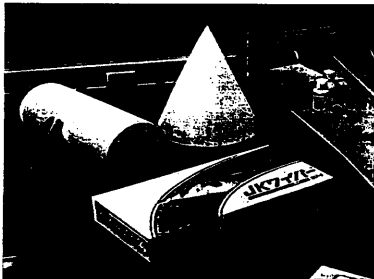


図16. 入力画像

(ティッシュペーパーの箱のあるシーン)

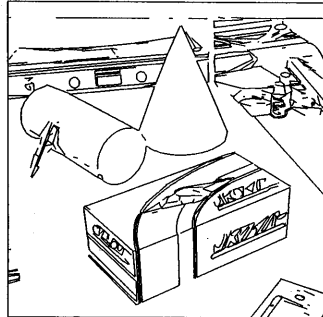


図17. 図16の2次元境界表現

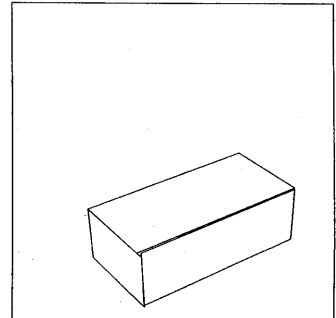


図18. 発見された

ティッシュペーパーの箱

を表す行列は直交条件が満たされなければならない、そのため何らかの手法が必要である。本章では回転成分の回転角度を直接計算する方法を紹介する。定量的照合の過程は3段階に分かれている。まず最初に局所的な特徴を使って変換の初期値が決められる。その後変換は残りの特徴を用いて正当性を検証され、最後に微調整される。以下の説明では、定性的照合によって画像から発見された図12の直方体を画像データとして、またそれに対するモデルである図19の直方体を用いて説明を行う。

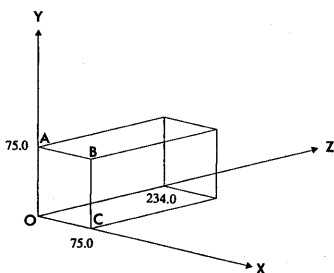


図19. 図3(a)の画像中の直方体の3次元モデル

4. 1 初期照合

モデルが与画像への3次元-2次元座標変換のパラメータを得るためには少なくとも三つの対応点を求める必要がある。モデル側の対応点としては例えば図19の点A、B、Cのような、連続した二つのセグメントの三つの端点を利用する。ここで (x_i, y_i, z_i) ($i=1, 2, 3$)をその3点の座標とする。また画像側の対応点として、例えば図12の画像のP、Q、Rを選び、これらの点の座標を $(X_i, Y_i, 1)$ ($i=1, 2, 3$)とする。最初に座標変換の平行移動成分がモデルの3点の重心と画像の3点の重心の間のベクトルを計算することによって得る。拡大/縮小と回転の成分は以下の方法で得られる。

点A、B、Cをその3点の重心からの相対座標で表して (x'_i, y'_i, z'_i) ($i=1, 2, 3$)とする。また同様に点P、Q、Rもそれらの重心からの相対座標で表して $(X'_i, Y'_i, 1)$ ($i=1, 2, 3$)とする。モデル側のz成分を無視して、点A、B、Cをそれぞれ点P、Q、Rに一致させるようなアフィン変換を考えると、次のようになる。

$$\begin{pmatrix} X'_i \\ Y'_i \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x'_i \\ y'_i \end{pmatrix} \quad (i=1, 2, 3)$$

図20に示すように (θ', ϕ') をモデルの点A、B、Cによって定められた面の法線ベクトルのオイラー角とする。するとアフィン変換は以下のような座標変換の合成として表すことができる。

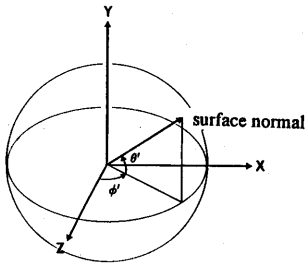


図20. 面の法線ベクトルのオイラー角表現

$$\begin{pmatrix} a_{11} & a_{12} & \cdot \\ a_{21} & a_{22} & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} = \sigma R_y(\phi) R_x(\theta) R_z(\alpha) R_x(-\theta') R_y(-\phi')$$

ここで σ は拡大/縮小を表し、 R_x 、 R_y 、 R_z はそれぞれx軸、y軸、z軸まわりの3×3の回転行列を表す。

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}$$

$$R_y(\phi) = \begin{pmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{pmatrix}$$

$$R_z(\alpha) = \begin{pmatrix} \cos \alpha & -\sin \alpha & 0 \\ \sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

上の式を解くことによって、 (ϕ, θ, α) と $(-\phi, -\theta, \alpha)$ の二つの回転角の組を得る。そして上記の変換と等価で、それぞれの軸の回りの回転がそれぞれただ一度だけとなるような、3次元座標変換の標準形式を導くことができる。

$$\begin{pmatrix} a_{11} & a_{12} & \cdot \\ a_{21} & a_{22} & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} = \sigma R_y(\pm\phi^*) R_x(\pm\theta^*) R_z(\alpha^*)$$

もし初期照合が正しければ、図21(a)(b)に示すようにこれら二つの解の一方は正しく、他方は誤っている。

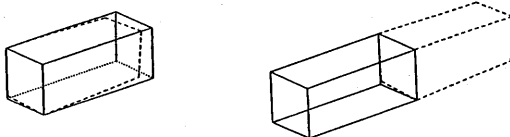


図21(a) 初期照合の結果(正解) 図21(b) 初期照合の結果(誤り)

4. 2 検証

初期照合で得られた座標変換の初期値が正しいかどうかを判定するために、モデルの他のセグメントを同じ初期値によって座標変換したときに、実際に存在している画像のセグメントにうまく一致するかどうかのテストを行う。

- ① 初期照合で得られたのと同じ座標変換を行ったモデルのセグメントのうち画像として見えるはずのものに対して、画像上で最も近い画像のセグメントを選ぶ(ρ - θ リストを使うことで容易に選ぶことができる)。
- ② 図22に示すように、モデルのセグメントの端点から画像のセグメントに対して垂線を下ろし、画像のセグメント上のその垂線の足を対応点とする。画像のセグメント上に垂線が下ろせなかった場合には端点を対応点とし、対応点に対するモデルのセグメント上の点を求める。以後これらをそれぞれモデル側の対応点、画像側の対応点と呼ぶことにする。
- ③ もしモデルと画像との誤差(誤差は対応点間の距離で定義する)に決められた値より大きいものがあれば、そのモデルのセグメントには対応する画像のセグメントが存在しないことになる。逆に誤差が小さい場合は、モデルのセグメントのうち画像のセグメントと一致した部分の長さ測定する。
- ④ モデルのセグメントの不一致の部分については再帰的に他の最も近い画像のセグメントを探し、同じ方法で一致した部分の長さが測定される。
- ⑤ モデルのセグメントの一致した部分の長さの合計、モデルの見える部分のセグメントの全長に対する比を、その物体が画像と一致している度合として計算する。一致した量が閾値を越えると、物体が発見されたと見なされる。逆に一致した量がその閾値を越えることができなければ、その座標変換は誤りだったと見なす。
- ⑥ もし初期照合で得られた二つの座標変換の候補のうち

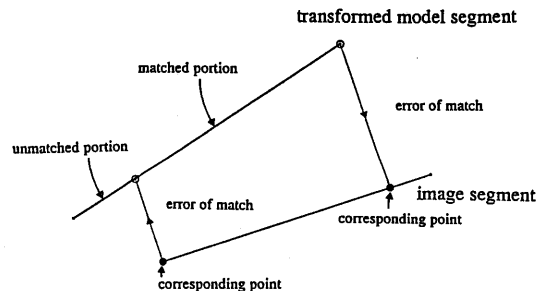


図22. 対応点の抽出

ち、一方を使って座標変換が誤りとなった場合、他方の候補を使って同様の対応を行う。第二の候補によっても変換が誤りとなった場合、最初の初期照合を誤りと見なして新しく別の初期照合を求め直す。

4.3 微調整

検証の結果、初期照合が正しいと判定された場合でも、それは局所的な特徴によって決められたものである。図21(a)に示すようにいくらかの誤差を含んでいる。したがって座標変換のパラメータは、検証のステップで新しく得られた対応点の誤差を使ってさらに正確になるように微調整される。その手順を以下に示す。

- ① G_n をモデル側の対応点すべての重心とする。また G_i を対応する画像側の対応点の重心とする。これより座標変換の平行移動成分を $G_i - G_n$ にする。
- ② モデルと画像とで対応が得られているセグメントのうち、 i 番目のものについて考える。初期照合の時と同じ様にモデル側の対応点3点を画像側の対応点3点に一致させるような座標変換 $\sigma_i R_x(\phi_i) R_y(\theta_i) R_z(\alpha_i)$ および $\sigma_i R_x(-\phi_i) R_y(-\theta_i) R_z(\alpha_i)$ を計算する。ここで用いられる3点は、モデル側および画像側の対応点(2点ある)と、 G_n および G_i を用いて計算される。この結果、正しい座標変換により近いものが選ばれる。もし新しく得られた座標変換のどちらも初期照合で得られた結果と大きく異なっていれば、それらは③のステップでは用いない。
- ③ ②のステップをすべての対応するセグメントについて行い、 $\bar{\sigma}$ 、 $\bar{\phi}$ 、 $\bar{\theta}$ 、および $\bar{\alpha}$ をそれぞれ σ_i 、 ϕ_i 、 θ_i 、および α_i の(画像セグメントの長さの重み付き)の平均値とする。この結果、新しい座標変換を $\bar{\sigma} R_x(\bar{\phi}) R_y(\bar{\theta}) R_z(\bar{\alpha})$ にする。
- ④ その新しい座標変換を用いて新しい対応点を求め、検証のステップを繰り返す。微調整は座標変換の変化が小さくなるまで繰り返される。

図23は4回の微調整の後の最終結果を示す。

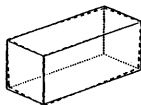


図23. 微調整の結果

4.4 実験結果

物体のモデルの大きさは既知であるので、単眼視で

はあるが σ の値から物体までの距離を知ることができる。 $f (= 1)$ をカメラの焦点距離とする。その時 f / σ はカメラから物体の重心までの距離を表す。表1は距離の測定の結果を示す。

表1. 距離測定の結果

real distances	results of measurement
500 mm	464 mm
750 mm	757 mm
1000 mm	965 mm

5. 結論

画像中のブロックのような物体を認識するための定性的および定量的照合を提案した。最初に、物体のモデルを呼び出し初期照合を導くための、手がかりとなるプリミティブがボトムアップまたはデータ駆動の方法で発見される。次に、トップダウンまたはモデル駆動の方法で詳細な部分が定量的に分析される。この手法は建物のように、平行四辺形や直方体をその要素に持つものであれば、同じ様に動く。またわれわれはすでに曲面を持った物体にもこの手法を拡張している。今後はさらに拡張して多くの物体についての「共通モデル」の自動作成を考えたい。

参考文献

- [1] Fishler, R. B., Model Invocation for Three Dimensional Scene Understanding, Proc. 10th IJCAI, 1987, pp. 805-807.
- [2] Burns, J. B. and Kitchen, L. J., Recognition in 2D Images of 3D Objects from Large Model Bases Using Prediction Hierarchies, Proc. 10th IJCAI, 1987, pp. 763-766.
- [3] 小泉, 富田: ソリッドモデルと画像の定量的照合のための定性的照合, 情報処理学会第36回全国大会
- [4] 富田, 小泉: ソリッドモデルと画像の照合, 情報処理学会第35回全国大会
- [5] Brooks, R. A., Symbolic Reasoning Among 3-D Models and 2-D Images, PhD dissertation, Stanford University, 1981.
- [6] 3次元物体の境界線に基づく記述の生成と認識, 信学会論文誌, Vol. J71-D No. 2, pp. 343-349, 1988
- [7] 富田, 高橋: 画像のB-REPのためのアルゴリズム, 信学会パターン認識と理解研究PRU86-87, 1987.
- [8] Roberts, L., "Machine Perception of Three-Dimensional Information," in Tippett, J. (Ed.), Optical and Electro-Optical Information Processing, pp. 159-197, MIT press, Cambridge MA., 1965.