

パラメータ空間分割による段階的モデルマッチング を用いた3次元物体認識

友納 正裕

油田 信一

筑波大学 知能ロボット研究室

〒305-8573 つくば市天王台 1-1-1

本稿では、1枚の画像から多面体モデルを用いて3次元物体を認識する手法を提案する。この手法は、モデルエッジと画像エッジを照合することにより物体認識と姿勢推定を行う。その際、透視投影での投影線分に関する不変量を用いて、6次元の姿勢空間を平行移動パラメータと回転パラメータの2つの3次元空間に分割し、計算量を削減する。この方法は、物体の姿勢範囲が限定されている場合に計算量を大幅に削減できる特徴がある。また、画像エッジは直線部分が抽出されていればよく、弱コントラストによる雑音エッジやオクルージョンによるエッジ分断に強い。

3-D Object Recognition based on Two-Staged Model Matching using a Projective Invariant

Masahiro TOMONO

Shin'ichi YUTA

Intelligent Robot Laboratory, University of Tsukuba

Tsukuba Ibaraki, 305-8573 JAPAN

{tomono,yuta}@roboken.esys.tsukuba.ac.jp

This paper presents a method for 3-D object recognition from a single 2-D image based on a polyhedral model. The method matches model edges with image edges to identify a target object and to estimate its pose. In the process, the method divides the six-dimensional pose space into two three-dimensional spaces, that is, location space and orientation space using an invariant with respect to perspective projection. This makes the time complexity low-order polynomial. Furthermore, the recognition time can be greatly reduced if the orientation is constrained because the complexity is proportional to the range of object's orientation. The method is robust against noise and small occlusion in image edges.

1 はじめに

3次元物体認識には様々な手法が提案されているが[3, 7, 6, 4, 1, 5]、実用化には多くの困難がある。とくに問題となるのは、計算量とロバスト性である。3次元物体認識は探索空間が大きく、計算量が膨大である。これは、物体認識の問題が、モデル特徴と画像特徴の対応づけ問題と物体姿勢を求める問題とを同時に解く必要があるためである。たとえば、Alignment法[4]では、モデル特徴点数を M 、画像特徴点数を N として、 $O(M^3N^3)$ ないし $O(M^2N^2)$ の計算量がかかる。Geometric hashing[6]も $O(N^4)$ の計算量がかかる。

ロバスト性は、とくに画像の下位レベル解析における特徴抽出で問題となる。たとえば、特徴ベースの認識手法では、エッジやコーナーなどの特徴を用いるが、これらの特徴は照明条件やコントラストにより抽出結果が変動し、種々の条件下で細かい特徴まで安定に抽出するのは難しい。しかも、オクルージョンがあると、本質的に特徴抽出できない部分が生じる。したがって、物体認識には、不完全な特徴抽出のもとでも認識できるロバスト性が要求される。

本稿では、3次元多面体をモデルとし、1枚の画像とモデルとでエッジ単位に照合を行うことにより、透視投影のもとで物体認識と姿勢推定を行う方法を提案する。エッジ照合は2段階から成り、まず画像エッジと画像平面に投影されたモデルエッジとが同一直線になりうるかどうかを調べ、次に両エッジの端点が一致するかどうか調べる。これを物体の各姿勢について行う。その際、透視投影での投影線分に関する不変量を用いて、6次元の姿勢空間を回転の3次元空間と平行移動の3次元空間に分割することで、計算量を削減する。 M をモデルエッジ数、 N を画像エッジ数とすると、計算量は $kM(N+N^3)$ となる。ただし、 N' はモデルエッジとの同一直線性で絞り込まれた画像エッジ数であり、 $N' \ll N$ であるため、実質的な計算量はかなり小さくなる。また、 k は物体のとりうる姿勢角度の範囲に比例する定数であり、姿勢角度範囲が限定できる場合は、 k の値は大幅に小さくできる。

本方式で用いるエッジは直線分であり、画像から安定して抽出できる部分でよい。このため、特徴抽出段階での雑音エッジに強い。また、特徴点としてエッジの端点を用いるが、画像から直接抽出するのではなく、画像エッジの直線方程式の交点として求める。このため、オクルージョンによって画像エッジが分断されても端点を求めることができる。以上により、特徴抽出の不完全さを吸収して、ロバスト性を高めることができる。さらに、透視投影のまま認識を行うため、従来の平行投影あるいは弱透視投影で近似する方式に比べ[4, 6, 5]、近似誤差の問題は生じない。

以下、2章で本方式の原理、3章でアルゴリズム、4章で実装について説明する。5章で実験結果を示し、6章で考察を述べる。

2 物体認識方式

本節では、物体モデルと画像間でエッジを照合することにより物体認識を行う方法を述べる。エッジ照

合の際に、カメラと物体の相対姿勢関係も同時に算出される。

2.1 準備

物体モデルとカメラの姿勢関係を図1に示す。物体モデルのローカル座標系を F 、カメラ座標系を F^c とする。

物体モデルの形状は、物体モデル座標系 F で定義される。物体モデル座標系 F からカメラ座標系 F^c への座標変換パラメータを $\tau = \langle \tau_r, \tau_t \rangle$ とする。 $\tau_r = (\theta, \phi, \psi)$ は回転成分(オイラー角)、 $\tau_t = (x^t, y^t, z^t)^T$ は平行移動ベクトルである。このとき、物体モデル上の点 P のカメラ座標系 F^c での値 P^c は次のようになる。

$$P^c = R(\tau_r)P + \tau_t \quad (1)$$

$R(\tau_r)$ は τ_r による回転行列である。

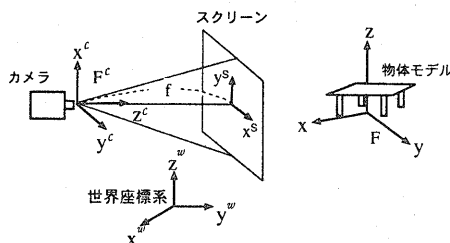


図1: カメラと物体モデルの姿勢関係

カメラ座標系における物体モデルの姿勢を用いて、ピンホールカメラモデルによる透視投影を行う。カメラ座標系の点 $P^c = (x^c, y^c, z^c)^T$ のスクリーン座標系での点 $P^s = (x^s, y^s)$ は以下になる。 P^s は τ の関数であることに注意。

$$P^s = (x^s, y^s)^T = \left(f \frac{y^c}{z^c}, f \frac{x^c}{z^c}\right)^T$$

2.2 エッジ対応に基づく物体認識

入力画像とモデルの照合は、入力画像中の2Dエッジ(画像エッジ)とモデルの3Dエッジ(モデルエッジ)とで1本ずつ対応をとることで行う。この節では、画像エッジの端点が完全に抽出された理想的な場合を考える。モデルは多面体なので、モデルエッジはすべて直線分である。画像エッジも、前処理によりすべて直線分に分割されているとする。

画像エッジの集合を L 、モデルエッジの集合を E とする。モデルエッジ $e \in E$ を姿勢 τ によって2次元画像に投影したエッジ(投影エッジ)を e^s あるいは $e^s(\tau)$ と表す。ここで、 P_1, P_2 を e の端点として、 e を $\langle P_1, P_2 \rangle$ と表す。同様に、 e^s を $\langle P_1^s, P_2^s \rangle$ 、 $l \in L$ を $\langle Q_1, Q_2 \rangle$ と表す。

入力画像から 3D モデルを認識する問題を、以下の式を 0 にする姿勢 τ と、 E から L への写像 $m = \{(e, l) \mid e \in E, l \in L\}$ を求める問題として定式化する¹。

$$S = \{(\tau, m) \mid \sum_{(e, l) \in m} D(e^s(\tau), l) = 0\}$$

D は 2 つの線分の距離であり、2 つの線分の両端点間の距離の 2 乗の和で定義する。すなわち、

$$D(e^s, l) = \min \{ \|P_1^s - Q_1\| + \|P_2^s - Q_2\|, \|P_1^s - Q_2\| + \|P_2^s - Q_1\| \}$$

姿勢 τ は、回転 $\tau_r = (\theta, \phi, \psi)$ と平行移動 $\tau_t = (x, y, z)^T$ の 6 変数から構成される。 τ_r の集合を T_r 、 τ_t の集合を T_t 、 m の集合を M とすると、 S の探索空間は $T_r \times T_t \times M$ となり、計算量は膨大となる。そこで、探索空間を分割し、解の計算を段階的に行うことで、計算量の軽減を図る。直観的には、まずエッジの直線方程式が一致しうかどうかで対応エッジを絞り込み、次に端点の一致により対応エッジを決定する。これを定式化すると以下ようになる。

$$S_1 = \{(\tau, m) \mid \sum_{(e, l) \in m} D_1(e^s(\tau), l) = 0\}$$

$$S_2 = \{(\tau, m) \mid \sum_{(e, l) \in m} D(e^s(\tau), l) = 0, (\tau, m) \in S_1\}$$

D_1 は、 e の投影エッジ e^s と l が端点を共有するときの両エッジの傾きの差である。

$$D_1(e^s(\tau), l) = |\text{slope}(e^s(\tau)) - \text{slope}(l)|, \text{ where } P_1^s = Q_1$$

ここで、 $\text{slope}(x)$ はエッジ x の傾き、 P_1^s はモデルエッジ e の端点 P_1 の投影点、 Q_1 は画像エッジ l の端点である。これより、 S_1 は直線方程式が一致しうモデルエッジ e と画像エッジ l の対の集合となる。

S_2 は、 S_1 で得られたエッジ対応候補のうち、2 つのエッジの両端点が一致するものの集合である。以下の命題が成り立つ。

命題 1 $S_2 = S$.

証明. S_2 の定義域は S の定義域の部分集合となるので、明らかに $S_2 \subset S$ である。次に、 l と e^s の両端点一致すれば傾きも一致するので、 $(\tau, m) \in S$ ならば $(\tau, m) \in S_1$ が成り立つ。したがって、 $(\tau, m) \in S_2$ となり、 $S_2 \supset S$ が成り立つ。□

これにより、 S を解くには、まず S_1 を求め、次に S_2 を求めればよい。さらに、次の補題により、 S_1 は τ_r にだけ依存することがわかる。

補題 1 $P_1^s = Q_1$ という制約を満たす τ_r, τ_t において、 e^s の傾きは τ_r だけに依存し、 τ_t に対しては不変である。

¹ m は部分写像である。対応する画像エッジをもたないモデルエッジがあってもよい。

証明. 点 $Q_1 = (u_1, v_1)^T$ と点 $P_1^s = (x_1^s, y_1^s)^T$ が一致するので、

$$(u_1, v_1)^T = (x_1^s, y_1^s)^T = (f \frac{y_1^c}{z_1^c}, f \frac{x_1^c}{z_1^c})^T \quad (2)$$

が成り立つ。これと、 $(x_2^s, y_2^s)^T = (f \frac{y_2^c}{z_2^c}, f \frac{x_2^c}{z_2^c})^T$ とから、姿勢 τ における投影エッジ e^s の傾きは以下のようになる。

$$\begin{aligned} \text{slope}(e^s) &= \frac{y_2^s - y_1^s}{x_2^s - x_1^s} = \frac{z_1^c(x_2^c - x_1^c) - x_1^c(z_2^c - z_1^c)}{z_1^c(y_2^c - y_1^c) - y_1^c(z_2^c - z_1^c)} \\ &= \frac{f(x_2^c - x_1^c) - v_1(z_2^c - z_1^c)}{f(y_2^c - y_1^c) - u_1(z_2^c - z_1^c)} \end{aligned} \quad (3)$$

ここで、式 (1) より、

$$P_2^c - P_1^c = R(\tau_r)(P_2 - P_1)$$

となり、 $P_2^c - P_1^c = (x_2^c - x_1^c, y_2^c - y_1^c, z_2^c - z_1^c)^T$ は、 τ_t によらない。また、 u_1, v_1 は画像エッジで決まる。よって、式 (3) は τ_t に対して不変である。□

補題 1 により、 S_1 は、 τ_r を変化させながら、 $\sum_{(e, l) \in m} D_1(e^s(\tau), l) = 0$ を満たす $m \in M$ を求めることで得られる。このとき、 m は τ_r の関数になる。

次に、 τ_t に関して以下の補題が成り立つ。

補題 2 $P_1^s = Q_1$ かつ $\text{slope}(e^s) = \text{slope}(l)$ を成り立たせる各 τ_r に対して、モデルエッジ $e \in E$ 上の端点 $P_2 (\neq P_1)$ に対して、 $P_2^s = Q_2$ となる τ_t がただ 1 つ存在する。

証明 付録参照。

ここで、

$$\tau_t = FR(\tau_r)(P_2 - P_1) - R(\tau_r)P_1 \quad (4)$$

$$F = \frac{1}{v_2 - v_1} \begin{pmatrix} v_1 & 0 & \frac{v_1 v_2}{f} \\ u_1 & 0 & \frac{u_1 v_2}{f} \\ f & 0 & v_2 \end{pmatrix} \quad (5)$$

である。ただし、 $Q_1 = (u_1, v_1)^T, Q_2 = (u_2, v_2)^T$ である。

補題 2 により、 S_1 で得られた各 τ_r に対して τ_t は式 (4) により算出できる。

以上より、 S を求めるには、まず画像エッジと投影エッジの直線方程式が一致しうかどうかで τ_r を絞り込み (S_1)、次に画像エッジと投影エッジの端点一致するかどうかで τ_t を求める (S_2)、というように探索空間を分割して段階的に行うことができる。

2.3 エッジ端点が検出できない場合の処理

前節ではエッジ端点が完全に抽出された理想的な状況を想定したが、実際の画像においてエッジ端点の抽出を完全に行うのは難しい。照明条件やコントラストによって、エッジが途中で切れたり、本来分離されるべきエッジがつながって長くなることもある。これ

らの問題により、実画像においてエッジ端点を特徴点として用いるのはロバスト性の観点からあまり現実的とは言えない。

この問題に対処するため、モデルにおいて仮想エッジを導入する。モデル形状を定義するエッジを実エッジと呼ぶとすると、仮想エッジは、実エッジと同一直線上にあって実エッジと重なる直線分である。仮想エッジの端点は他の仮想エッジとの交点として定義する。すると、画像に投影した仮想エッジの端点は、投影エッジ同士の交点として求めることができる。

エッジの直線部分は比較的安定して抽出できるため、投影エッジの交点を求めることは難しくなく、誤差も比較的小さい。しかも、オクルージョンがあっても、直線部分がある程度抽出できていれば、交点を求めることができる。なお、これ以降、仮想エッジとモデルエッジを同じ意味で扱う。

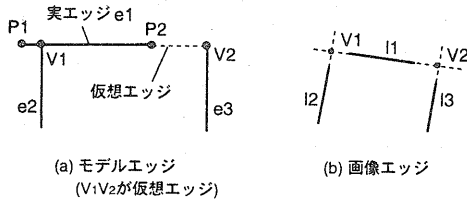


図 2: 仮想エッジ

前節で述べたように S_1 の計算には画像エッジの端点 Q_1 を必要とする。しかし、もし上記のように画像エッジが不完全であるとして、真の端点を画像エッジの交点として求めるとすると、 S_1 の計算量は $|L|^3$ に比例することになる。しかし、次の補題により、端点 Q_1 の代わりに画像エッジ l 上の任意の点を用いても S_1 は変わらないことが示される。これにより、 S_1 は l の端点(真の端点ではない)を用いて計算してよく、その計算量は $|L|$ に比例する。

まず、前節の D_1 の定義を少し変える。

$$D_1(e^s(\tau), l) = |\text{slope}(e^s(\tau)) - \text{slope}(l)|, \text{ where } P_1^s = Q$$

Q は l 上の任意の点である。次の補題により、この変更は S_1 に影響を与えない。

補題 3 姿勢 $\tau = \langle \tau_r, \tau_t \rangle$ において、 $P_1^s = Q$ かつ $\text{slope}(e^s) = \text{slope}(l)$ が成り立つとする。このとき、 $D_1(e^s(\tau), l) = 0$ を満たす τ_r の値は Q に依存しない。

証明. l 上の任意の2点 $Q = (u, v)$, $Q' = (u', v')$ に対して、 $P_1^s = Q$ および $\text{slope}(e^s) = \text{slope}(l)$ を満たす姿勢を $\tau = \langle \tau_r, \tau_t \rangle$, $P_1^s = Q'$ および $\text{slope}(e^s) = \text{slope}(l)$ を満たす姿勢を $\tau' = \langle \tau'_r, \tau'_t \rangle$ とする。

$\text{slope}(l) = k$ として、補題 1 より、

$$k = \frac{f dx^c - v dz^c}{f dy^c - u dz^c} = \frac{f dx^{c'} - v' dz^{c'}}{f dy^{c'} - u' dz^{c'}} \quad (6)$$

ただし、 $(dx^c, dy^c, dz^c)^T = (x_2^c - x_1^c, y_2^c - y_1^c, z_2^c - z_1^c)^T$, $(dx^{c'}, dy^{c'}, dz^{c'})^T = (x_2^{c'} - x_1^{c'}, y_2^{c'} - y_1^{c'}, z_2^{c'} - z_1^{c'})^T$ であり、 $(x_i^{c'}, y_i^{c'}, z_i^{c'})^T$ は、 τ' による点 $(x_i, y_i, z_i)^T$ の式 (1) の結果である。

式 (6) を変形して、以下を得る。

$$\begin{aligned} f dx^c - k f dy^c + (ku - v) dz^c &= 0 \\ f dx^{c'} - k f dy^{c'} + (ku' - v') dz^{c'} &= 0 \end{aligned} \quad (7)$$

ここで、 l の直線方程式を $\alpha(x - u_1) + \beta(y - v_1) = 0$ とすると、 $k = -\frac{\alpha}{\beta}$ となり、 Q' は l 上にあるので、 $u' = u + \beta t$, $v' = v - \alpha t$ と表せる。これより、 $ku' - v' = k(u + \beta t) - (v - \alpha t) = ku - v$ となり、これを用いて、式 (7) を変形すると、以下を得る。

$$\begin{aligned} (f, -kf, ku - v) R(\tau_r) (dx, dy, dz)^T &= 0 \\ (f, -kf, ku - v) R(\tau'_r) (dx, dy, dz)^T &= 0 \end{aligned}$$

ただし、 $(dx, dy, dz)^T = (x_2 - x_1, y_2 - y_1, z_2 - z_1)^T$ である。この2つの式は τ_r, τ'_r に関する同じ方程式であり、 Q と Q' のどちらに対しても、 τ_r の解の集合は同じになる。以上より、 τ_r の値は Q に依存しない。□

3 アルゴリズム

エッジ照合の解 S は、 S_1 と S_2 、エッジ照合の3段階に分けて求める。それぞれに対する解探索アルゴリズムを示す。

3.1 S_1 の探索

S_1 の探索は、各 τ_r に対して、 $D_1(e^s(\tau), l) = 0$ を満たすエッジ対応 m を求めることで行う。 τ_r は角度であるので $0 \sim 360$ 度で閉じており、適当に量子化して網羅的に探索することができる。ただし、実画像は誤差を含むため、 $D_1(e^s(\tau), l) = 0$ という条件が成り立つとは限らない。そこで、適当な閾値により直線方程式が一致しているかを判断し、 S_1 の解の候補とする。

以下にアルゴリズムを示す。

- (1) $\tau_r = (\theta, \phi, \psi)$ の定義域を適当な区画で離散化する。各区画には以下で計算するスコアが記録される。
- (2) τ_r の各離散値に対して $G_1(\tau_r)$ を計算し、スコアとする。 $G_1(\tau_r)$ は、 τ_r において対応する画像エッジをもつモデルエッジの本数である。

$$G_1(\tau_r) = \sum_{e \in E} c \left(\bigvee_{l \in L} (D_1(e^s(\tau), l) \leq th_1) \right)$$

$$c(a) = \begin{cases} 1 & \text{if } a \text{ is true} \\ 0 & \text{if } a \text{ is false} \end{cases}$$

th_1 は閾値であり、 $D_1(e^s(\tau), l) < th_1$ であれば、 $D_1(e^s(\tau), l) = 0$ が成り立っていると見なす。

ここで、 $D_1(e^s(\tau), l) \leq th_1$ となる (e, l) を m に加える。 m は τ_r の関数となるので $m(\tau_r)$ と表す。 $m(\tau_r)$ は次のように表せる。

$$m(\tau_r) = \{(e, l) \mid (e, l) \in E \times L, D_1(e^s(\tau), l) \leq th_1\}$$

- (3) 得点の多い $(\tau_r, m(\tau_r))$ を解の候補とする。

$$S'_1 = \{(\tau_r, m(\tau_r)) \mid G_1(\tau_r) \geq th_2\}$$

th_2 は適当な閾値である。

3.2 S_2 の探索

S_2 は、viewpoint consistency constraint [7] に基づいて、投票 (voting) によって求める。この制約は、画像に投影されるモデル特徴の位置は、1 つの視点からの投影に関して整合がとれていなければならないというものである。そこで、各エッジの端点から式 (4) により τ_t を求め、 τ_t に関する投票数を整合性の度合と見なす。すなわち、最も得票数の多い τ_t および τ_r が最も整合性が高いとする。

- (1) τ_t の範囲を適当な区画で分割し、投票用の bin を作る。 τ_t の範囲は無限であるため原理的には bin の個数は無限大になるが、ここでは、適当な区間に範囲を限定し、この区間で bin を作る。
- (2) S'_1 の各 $(\tau_r, m(\tau_r))$ について、以下の処理を繰り返す。
- (3) 各 $(e, l) \in m(\tau_r)$ に対して、以下のように端点候補を計算する。
 - (a) モデルエッジ e の両端点 P_1, P_2 で e と接続するモデルエッジ e_1, e_2 を求める。 e に対する e_1, e_2 はあらかじめモデル定義で与えておく。
 - (b) e_1, e_2 に対応する画像エッジ群を $m(\tau_r)$ から求め、それぞれ L_1, L_2 とする。
 - (c) l と $l_1 \in L_1$ の交点 Q_1 、および、 l と $l_2 \in L_2$ の交点 Q_2 を求める。 Q_1, Q_2 が画像エッジ l の端点候補となる。
- (4) エッジの組み合わせ $(e, l) \in m(\tau_r)$ に対して、(3) で求めた端点 Q_1, Q_2 を用いて、式 (4) より τ_t を計算する。そして、得られた τ_t に対応する bin に 1 票加える。
- (5) (4) で得た bin の得票数 $votes(\tau_t)$ が閾値 th_3 を越えれば、その (τ, m) を解の候補とする。

$$S'_2 = \{(\tau, m(\tau_r)) \mid (\tau_r, m(\tau_r)) \in S'_1, votes(\tau_t) \geq th_3\}$$

ただし、3 次元空間で τ_t の投票を行うと精度が悪くなるので、実用的には z^t の 1 次元で投票を行う。

3.3 エッジの対応づけ

S'_2 で得られた各 $(\tau, m(\tau_r))$ に対して、実際に最もよく一致する投影エッジと画像エッジの組合せを求めて候補を絞り込み、解を確定する。処理を以下に示す。

- (1) 各 $(\tau, m(\tau_r)) \in S'_2$ に対して、(2)、(3) を行う。

- (2) 同一直線性によるフィルタリング

エッジ対 $(e, l) \in m(\tau_r)$ において、 l が e^s と傾きがほぼ同じで、しかも、 e^s の近傍に含まれるかどうかを調べる (図 3)。この条件を満たさなければ、この (e, l) を $m(\tau_r)$ から除く。

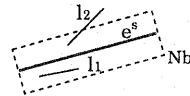
- (3) 端点距離によるスコアリング

(2) で残った (e, l) に対して、 $D(e^s, l)$ が最小となる l を e^s に対応するエッジであるとする。また、このときの $D(e^s, l)$ を e のスコアとする。もし、 e が対応する画像エッジを持たなければ、 e に適当なペナルティ点を与える。

- (4) スコアによる姿勢の決定

$\sum_{e \in E} score(e, \tau)$ が最小となる τ を解とする。ただし、 $score(e, \tau)$ は e に対する最小の $D(e^s(\tau), l)$ かペナルティ点である。

ペナルティ点を与えるのは、画像エッジにマッチするモデルエッジがない $m(\tau_r)$ が解として選ばれるのを防ぐためである。



Nb は投影エッジ e^s の近傍範囲。このフィルタにより、 l_1 は残り、 l_2 は除外される。

図 3: 同一直線フィルタ

3.4 計算量

S'_1 の計算量は以下になる。

$$C_1 = N_\theta N_\phi N_\psi |E| |L| \quad (8)$$

N_θ, N_ϕ, N_ψ は、それぞれ、 θ, ϕ, ψ の定義域の区画分割数である。物体姿勢の範囲に制約があれば N_θ, N_ϕ, N_ψ は小さくなり、その結果 C_1 も小さくなる。

S'_2 の計算量は以下になる。

$$C_2 = |S'_1| Avg_1(|E'|) Avg_2(|L'|)^3 \quad (9)$$

E' は、モデルエッジのうち、 S'_1 の $m(\tau_r)$ に含まれるものの集合である。 Avg_1 は、 τ_r 上で平均をとる関数である。 L' は S'_1 の $m(\tau_r)$ において、1 本のモデルエッジに対応する画像エッジの集合である。 Avg_2 は、 τ_r および E' 上で平均をとる関数である。 C_2 は $Avg_2(|L'|)$ の 3 乗で効くが、 L' は e^s と同一直線である画像エッジに絞り込まれているので、 L に比べるとかなり小さくなる。なお、画像エッジの端点が交点を用いずに直接求められる場合の C_2 は、

$$C_2 = |S'_1| Avg_1(|E'|) Avg_2(|L'|)$$

となる。

また、エッジ対応の計算量は以下になる。

$$C_3 = |S'_2| Avg_1(|E'|) Avg_2(|L'|)^3 \quad (10)$$

4 実装

本方式により物体認識を行うプログラムを実装した。処理の流れを図4に示す。まず Edge detector により、Canny オペレータ [2] を用いて、適当なエッジ抽出閾値でエッジを抽出する。次に、Edge cutter により、曲率の大きい点でエッジを直線分に分割する。この際、あらかじめ設定した閾値より小さい長さの直線分は除去される。次に、POI calculator は、得られた直線分 (画像エッジ) の交点を計算する。このとき、あらかじめ定めた範囲閾値よりも画像エッジの端点から離れた交点は除外する。

次に、Pose estimator により、上記画像エッジとその交点およびモデルエッジを用いて S'_1 および S'_2 を計算する。この際、 τ_r および τ_t の離散化幅を指定する。また、 τ_r の離散化による誤差に対処するため、投影エッジの傾きと画像エッジの傾きが一致したと判断する許容範囲を設定する。最後に、得られた τ の各候補に対して、Edge matcher により投影エッジと画像エッジの対応スコアを計算して、最終的な物体姿勢と対応エッジを求める。

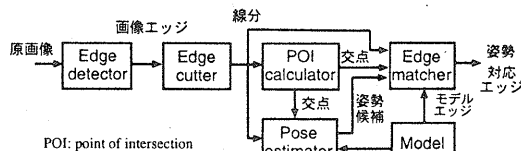


図 4: 処理の流れ

設定すべきパラメータを表1に示す。これらのパラメータ値はロバスト性の1つの指標になる。すなわち、パラメータの適切な値が環境やモデルなどの条件に依存する度合いが小さいほど、システムがロバストであると言える。たとえば、エッジ抽出閾値は照明やコントラストによって適切な値が変動し、ロバスト性を低くする要因となりうる。その値が小さすぎると、Edge detector は大量の雑音エッジを生成することがある。本手法は、その場合でもうまく対処できることが次節で示される。エッジ長閾値と交点範囲閾値は、本手法による認識に大きな影響を与えるが、その分析は次節以降で述べる。他のパラメータは、処理の精度や時間に影響するが、環境やモデルへの依存度はそれほど高くない。このため、次節で述べる実験では、これらのパラメータ値は固定とした。

5 実験

実験は、研究室の周辺にある物体をデジタルカメラで撮影して行った。画像は320×240に縮小して使った。マシンはパーソナルコンピュータ (CPU:PentiumII 400MHz、メモリ 96MB)、OS はLinux、言語はJava

表 1: 主要パラメータ

パラメータ	役割
エッジ抽出閾値	Canny オペレータの閾値
曲率閾値	エッジ分割点の曲率
エッジ長閾値	短いエッジを除去
交点範囲閾値	エッジの実端点と交点の距離
τ_r 離散化幅	S'_1 計算
τ_t 離散化幅	S'_2 計算
エッジ傾き許容範囲	e° と l の傾き一致の範囲
エッジ対応閾値	同一線分とみなせる範囲

(IBM JDK1.1.8) である。

対象とした物体は、液晶ディスプレイ、ドア、机、電話台、階段である。モデルは3次元ワイヤーフレームで構成した。モデルは認識に有用なエッジだけを持ち、必ずしも物体の形状を完全に表現するわけではない。たとえば、机は、通常の視点からはおもに前面部分しか見えないため、机モデルは前面部分のエッジだけから構成されるものに簡略化している。

液晶ディスプレイの認識結果を図5に示す。図左上が原画像である。原画像内の左下に使用したモデルを示した。図右上はエッジ抽出結果である。図左下はエッジ分割の結果であり、抽出された画像エッジのうち短いものは削除され、この例では62本が残っている。図右下は最終的な認識結果であり、得られた τ を用いてモデルを原画像に投影している。ティッシュペーパーの箱によるオクルージョンがあるが、正しく認識されているのがわかる。

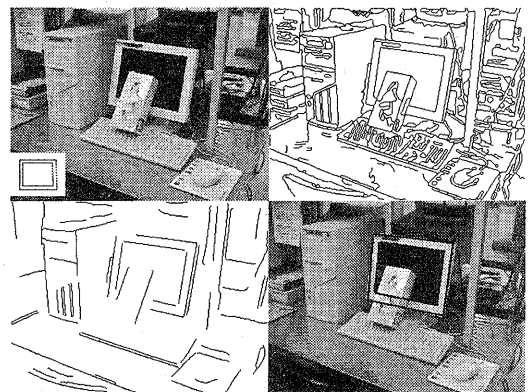


図 5: 液晶ディスプレイの認識結果

図6(a)は液晶ディスプレイの認識における S'_1 の分布である。横軸と縦軸は、それぞれ τ_r の ψ と ϕ である。 ψ は0~360度、 ϕ は0~180度の区間を離散化幅5度で区切ってある。対象物体はすべて水平面に置かれ

ているので、 $\theta = 0$ と固定している。この図では、各区画ごとに3.1節で述べた $G_1(\tau_r)$ の値を輝度で表している。明るい方が $G_1(\tau_r)$ の値が大きい。この実験ではモデルエッジ数が8本なので、最も明るい区画の $G_1(\tau_r)$ の値は8である。同図 (b) は S'_2 の分布である。この図では、 τ_r の各区画ごとに、その τ_r での $votes(z^t)$ の最大値を輝度で表している。 $\psi = 330, \phi = 30$ 付近で高い値が得られている。なお、 τ_t の離散化幅は10cmとしている。

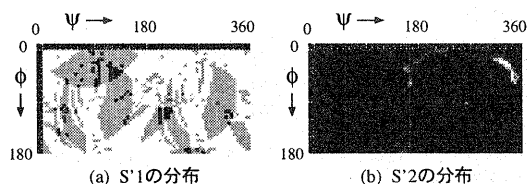


図 6: S'_1 と S'_2 の分布

処理のロバスト性を調べるために、エッジ抽出閾値を低くして雑音エッジを大量に生成した場合の実験を行った。その結果を図7に示す。ターゲットはドアであり、図左上が原画像、右上がエッジ抽出結果である。エッジ抽出閾値が低いため、雑音エッジが大量に発生している。図左下が線分抽出結果である。エッジ長閾値を長めにすることで、雑音エッジが除かれている。ドアの主特徴エッジである外枠と窓は比較的に長いので、除かれずに残っている。図右下が認識結果であり、消火器による小さなオクルージョンもあるが、選別された長いエッジを用いてドアが正しく認識されている。

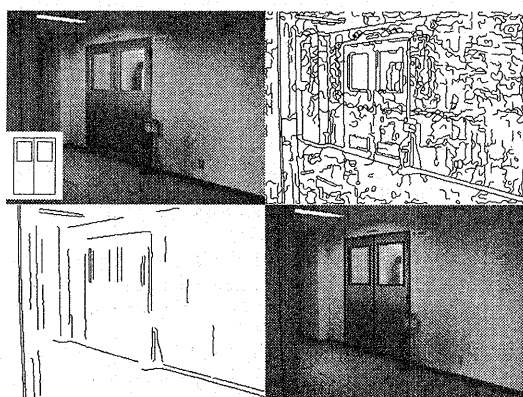


図 7: 雑音化でのドアの認識

机、電話台、階段の認識結果を図8に示す。用いたモデルを各画像内に小さな枠で示してある。各画像は、

得られた τ に基づいて、モデルを原画像に投影したものである。図右下は階段の認識失敗例である。

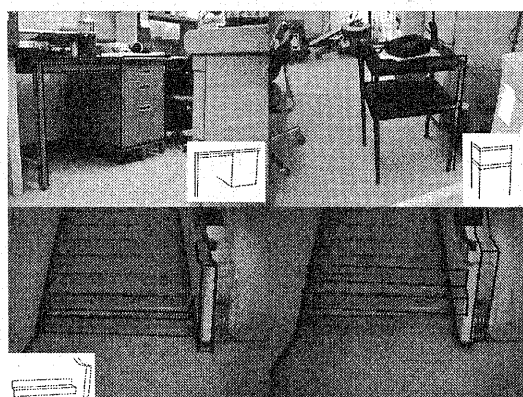


図 8: 机、電話台、階段の認識結果

処理時間はモデルや画像、さらに、前処理で得られた線分数や交点数によって異なるが、前処理は2~4秒、 S'_1 と S'_2 は3~20秒、エッジ対応検証は3~60秒である。エッジ対応検証が遅いのは、3.3節で述べた同一直線性検査のアルゴリズムの実装がアドホックなためであり、改善可能である。なお、3.4節の計算量の議論からわかるように、 S'_1, S'_2 の計算は、物体姿勢の事前知識があれば大幅に速くなりうる。

認識成功率の評価はまだ行っていないが、傾向として、液晶ディスプレイやドアは比較的安定して成功するが、机、電話台、階段はパラメータ値により失敗することもある。失敗の原因は、主特徴となる線分がうまく抽出されないためと考えられる。これについては、次節で述べる。

6 考察

実験の結果、認識結果に影響を与えるパラメータは、エッジ長閾値と交点範囲閾値であることがわかった。実験では、画像エッジ数や交点数が多いと誤認識(画像中の違う物体とマッチする)が増える傾向にある。エッジ長閾値が小さいと画像エッジ数が増え、また、交点範囲閾値が大きいと交点が増える。そのため、モデルエッジと照合する画像エッジ数が増えることになり、画像エッジのランダムな組合せがモデルと偶然一致する確率が増え、誤認識が起りやすくなると考えられる。この問題は、形状が単純でモデルエッジ数が少ない物体の場合にも現れる。これは、本手法のようなエッジベース方式の本質的な問題である。

エッジ長閾値の適切な設定は簡単ではない。長いエッジだけで主特徴が構成される物体の場合は、単にエッジ長閾値を大きくすればよい。実験例の液晶ディスプレイやドアはその例である。しかし、短いエッジ

が認識に決定的な役割を果たす場合がある。たとえば、実験例で示した階段では、各ステップの縦エッジは、長い横エッジの交点を決めるために必要であり、この縦エッジが除かれた場合は S'_2 を正しく計算できなくなる。このことから、エッジ長閾値をむやみに大きくすることはできず、他の方法で画像エッジ数を減らす必要がある。

これに対処する方法として、同一直線性と色の利用が考えられる。すなわち、同一直線上にある近い画像エッジを1つにまとめて、画像エッジ数を減らすと同時に1つのエッジ長を長くする。これにより交点範囲閾値を小さくしても、 S'_2 計算に有効な交点を得られる確率が増える。このとき、画像エッジの近傍の色を用いて同一物体の画像エッジかどうかを調べることが有効であろう。また、モデルの色と近い色をもつ画像エッジだけをあらかじめフィルタリングして、画像エッジ数を減らすことも有効であろう。

7 まとめ

多面体モデルを用いたエッジ照合による物体認識方式を提案した。本手法は、計算量が小さく、また、エッジ直線部分を用いるためロバスト性が高いという特長をもつ。

付録

補題2の証明

$P_1^s = Q_1, P_2^s = Q_2$ より、

$$(u_1, v_1)^T = (x_1^s, y_1^s)^T = (f \frac{y_1^c}{z_1^c}, f \frac{x_1^c}{z_1^c})^T$$

$$(u_2, v_2)^T = (x_2^s, y_2^s)^T = (f \frac{y_2^c}{z_2^c}, f \frac{x_2^c}{z_2^c})^T$$

となる。式(1)より、

$$P_1^c = R(\tau_r)P_1 + \tau_t$$

$$P_2^c = R(\tau_r)P_2 + \tau_t$$

である。これより、

$$\begin{aligned} \tau_t &= P_1^c - P_1^r = (\frac{v_1}{f}, \frac{u_1}{f}, 1)^T z_1^c - P_1^r \\ \tau_t &= P_2^c - P_2^r = (\frac{v_2}{f}, \frac{u_2}{f}, 1)^T z_2^c - P_2^r \quad (11) \end{aligned}$$

となる。ただし、 $R(\tau_r)P_1 = P_1^r, R(\tau_r)P_2 = P_2^r$ としている。ここで、

$$A = (a_1, a_2, a_3)^T = (\frac{v_1}{f}, \frac{u_1}{f}, 1)^T$$

$$B = (b_1, b_2, b_3)^T = (\frac{v_2}{f}, \frac{u_2}{f}, 1)^T$$

とおき、また、

$$P_1^r = (x_1^r, y_1^r, z_1^r)^T$$

$$P_2^r = (x_2^r, y_2^r, z_2^r)^T$$

として、式(11)を整理すると、

$$a_1 z_1^c - b_1 z_2^c + (x_2^r - x_1^r) = 0$$

$$a_2 z_1^c - b_2 z_2^c + (y_2^r - y_1^r) = 0$$

$$a_3 z_1^c - b_3 z_2^c + (z_2^r - z_1^r) = 0$$

となる。

これを z_1^c, z_2^c に関する連立方程式と見なすと以下のようになる。

$$H(z_1^c, z_2^c, 1)^T = 0 \quad (12)$$

ただし、

$$H = \begin{pmatrix} a_1 & -b_1 & x_2^r - x_1^r \\ a_2 & -b_2 & y_2^r - y_1^r \\ a_3 & -b_3 & z_2^r - z_1^r \end{pmatrix}$$

である。式(12)が解をもつためには、 $|H| = 0$ が必要十分条件である。

$$\begin{aligned} |H| &= -(A \times B) \cdot (P_2^r - P_1^r) \\ &= -\frac{1}{z_1^c z_2^c} ((P_2^r + \tau_t) \times (P_1^r + \tau_t)) \cdot (P_2^r - P_1^r) \end{aligned}$$

ここで、 $(P_1^r + \tau_t), (P_2^r + \tau_t), (P_2^r - P_1^r)$ は同一平面上にある3次元ベクトルである。 $(P_2^r + \tau_t) \times (P_1^r + \tau_t)$ はその平面に垂直になるから、 $(P_2^r - P_1^r)$ にも垂直になり、 $|H| = 0$ となる。よって、式(12)は解をもつ。

次に、式(12)を解くと、

$$\tau_t = FR(\tau_r)(P_2 - P_1) - R(\tau_r)P_1$$

となる。 F の値は式(5)である。これより、 τ_r を決めれば、 τ_t の値は一意に決まる。□

参考文献

- [1] R. Basri and E. Rivlin, Homing using Combinations of Model Views, *Proc. of IJCAI'93*, pp. 1586-1591, (Aug. 1993).
- [2] J. Canny, A Computational Approach to Edge Detection, *IEEE Trans. PAMI*, Vol. 8, No. 6, pp. 679-698, (Nov. 1986).
- [3] M. A. Fischler and R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications fo ACM*, Vol. 24, No. 6, pp. 381-395, (1981).
- [4] D. P. Huttenlocher and S. Ullman, Recognizing Solid Objects by Alignment with an Image, *International Journal of Computer Vision*, Vol. 5, No. 2, pp. 195-212, (1990).
- [5] D. W. Jacobs, Matching 3-D Models to 2-D Images, *International Journal of Computer Vision*, Vol. 21, No. 1/2, pp. 123-153, (1997).
- [6] Y. Lamdan, J. T. Schwartz, and H. J. Wolfson, On Recognition of 3-D Objects from 2-D Images, *Proc. of ICRA '88*, pp. 1407-1413, (1988).
- [7] D. G. Lowe, Three-Dimensional Object Recognition from Single Two-Dimensional Images, *Artificial Intelligence*, Vol. 31, pp. 355-395, (1987).