解説

三次元物体のモデリング・認識とロボティクスへの応用

3-D Object Modeling and Recognition for Robotic Tasks

友 納 正 裕^{*1} 山 崎 公 俊^{*2} ^{*1}千葉工業大学 ^{*2}東京大学

Masahiro Tomono^{*1} and Kimitoshi Yamazaki^{*2} ^{*1}Chiba Institute of Technology ^{*2}The University of Tokyo

1. は じ め に

人間の生活環境で作業を行うロボットにとって、物体をベースにした環境認識は極めて重要なテーマである。生活環境のほとんどは人工の物体に囲まれているため、あらゆる作業で物体認識が必要になると言っても過言ではない。物体の操作や運搬には対象物体の認識が必要であり、さらに、その作業場所や収納場所も机や棚など他の物体であることが多い。また、屋内を移動する場合、物体を目印（ランドマーク）として利用すると有用なことが多い。

物体を認識するには物体のモデルが必要である。物体モデルに基づいたナビゲーションや物体操作は、従来から研究されている[1]~[5]。だが、その多くはモデルを手で作成し、アドホックな認識手法を用いていた。また、モデルを手で作成するには多大な労力がかかるため、ロボットが物体モデルを自動的に取得することが望まれる。近年、プロセッサの飛躍的進歩とコンピュータビジョン技術の進展により、三次元でのモデル生成と認識を行うことが可能になってきた。

筆者らは、このような動向を踏まえて、物体をベースとした環境認識について研究してきた。本稿では、筆者らの研究を中心に、物体モデリング・認識の手法をナビゲーション、物体操作、地図構築に応用した研究事例を紹介する。

2. 物体モデリング

三次元物体モデルの作成には多大な労力がかかるため、センサから得たデータから（半）自動で物体モデルを生成する方法が研究されてきた。センサとして、レーザスキャナとカメラがよく用いられる。一般に、レーザスキャナは精度に関して有利であり、カメラは認識特徴の得やすさや色彩も含めたりアルさの観点で有利である。本稿では、物体認識を重視しているため、カメラを用いた方法を概説する。

2.1 物体形状とカメラ運動の同時復元

三次元物体モデルを生成するには、異なる視点からの複数の画像が必要である。そのためには、ステレオカメラと単眼カメラが利用可能であるが、本稿では、より一般性の高い単眼カメラを用いた方法を紹介する。単眼カメラで撮影した画像列から物体モデルを生成するには、物体形状とカメラ運動の同時復元（Structure from motion, SFM）を行う必要がある。

SFMは、画像間で対応づけられた特徴をもとに行われる。この特徴の抽出と対応づけには、例えば、KLTトラックが用いられる[6]。KLTトラックは、画像のコーナ点などの特徴点を抽出し、特徴点の近傍領域の類似性に基づいて画像間で特徴点を追跡する。SFMの方法としては、8点法[7]、因子分解法[8]、トリフォーカルテンソル[9]など多くの手法が提案されている。

山崎らは、因子分解法を用いて物体モデリングを行った[10]。因子分解法では、カメラモデルを線形近似することで、奥行の小さな物体に対して高い安定性を保つことができる。因子分解法で得られる復元結果は近似解であるが、これを初期値として非線形最小化を行うことにより透視投影の解が得られる。このとき、特徴点の誤対応が問題になるため、RANSAC[11]を用いてロバストな推定を行う。因子分解法は、特徴点の隠れ処理が煩雑になるため、初期段階のみ行い、以後はカメラ姿勢と物体形状を逐次的に復元する（本稿では、位置と方向を合せて姿勢と呼ぶ）。すなわち、時刻 $t-1$ の物体モデルを時刻 t の画像に投影して、その投影誤差が最小になるようにカメラの移動量を求め、そのカメラ移動量に基づいて、特徴点の三次元復元を行うことを繰り返す。

図1(a)は、SFMによる復元結果の例である。中央の点群がサッカーボールの復元形状であり、それを取り囲む円錐が、カメラの姿勢列を表している。ここでは、単眼カメラを搭載した車輪型ロボットに、物体の周囲を円弧走行させ、百数十枚の連続画像から復元を行った。

2.2 密な物体形状モデル

SFMでは、特徴点だけを三次元復元するので、疎な形状しか得られない。ナビゲーションや物体把持などの動作を

原稿受付 2008年1月30日

キーワード: Object Modeling, Object Recognition, Mobile Robot Navigation, Object Manipulation, Map Building

^{*1}〒275-0016 習志野市津田沼 2-17-1

^{*2}〒113-8656 文京区本郷 7-3-1

^{*1}Narashino-shi, Chiba

^{*2}Bunkyo-ku, Tokyo

図1 物体モデリングの例 (a) サッカーボールの特徴点復元とカメラ軌跡, (b) サッカーボールの密な復元, (c) 模型自動車の密な復元, (d) 机のエッジ復元とカメラ軌跡, (e) 机の密な復元

行うには、物体と自由空間の境界を知る必要があるため、ロボットにとって面情報を含んだ密な物体形状は重要である。

山崎らは、SFMによる特徴点の三次元復元に三角パッチを貼って面を復元し、さらにスペースカービング[12]を用いて詳細形状の復元を行った[10]。また、友納は、テクスチャの少ない物体に対処するため、まずエッジを復元し(図1(d))、さらにエッジ間に多数の線分を張り巡らせて面を復元した[13]。両方法とも、最終的には、物体形状をボクセル(三次元の微小区画)の集合に変換する。ボクセル表現は、三次元空間の占有格子地図と同等であり、ロボットの動作計画に用いるのに適している。図1(b), (c), (e)に、密な三次元モデルの例を示す。

3. 物 体 認 識

物体認識とは、入力画像に写っている物体を物体データベースから見つけることである。必要に応じて、物体の画像上の位置や三次元空間内の姿勢を求める処理も含む。ロボットへの応用では、物体の三次元空間での姿勢を知ることが重要である。友納は、SIFT記述子[14]と三次元形状を統合した物体モデルを用いて物体認識を実現した[15]。以下にその概要を述べる。

物体モデルは、三次元モデルと二次元モデルからなる。テクスチャの少ない物体にも対処できるように、特徴としてエッジ点を用いる。三次元モデルは、図1(d)と同様に、画像列からSFMにより復元した三次元エッジ点で構成する[13]。二次元モデルは、その画像列から抜粋した画像のエッジ点で構成する。エッジ点には、認識用の特徴ベクトルとしてSIFT記述子を付加する。

物体認識では、二次元モデルを用いて物体検出を行い、得られた候補に対して三次元モデルを用いて姿勢推定を行う。物体検出は、SIFT記述子を用いて入力画像とモデル画像をマッチングすることで行う。その際、全モデル画像のSIFT記述子からなるkd木を用いて[14]、入力エッジ点と類似したSIFT記述子をもつモデルエッジ点を高速に検

図2 物体認識の例 上段：認識用モデル，下段：認識結果

索する。物体形状による幾何学的な整合性チェックで誤対応を取り除いた後、類似したエッジ点を多くもつモデルを認識候補とする。姿勢推定では、三次元モデルを入力画像に投影したときに、その投影像と画像中の対象物体との誤差が最小となるカメラ姿勢を計算する。

図2に物体認識の例を示す。図2では、三次元モデル(白色)を入力画像に重ね合せている。エッジ点を用いているため、テクスチャの少ない物体でも、テクスチャが複雑な物体でも認識でき、また、オクルージョンや照明変化にもロバストであることが確認されている[15]。

4. 物体認識のナビゲーションへの応用

従来、移動ロボットのナビゲーションでは、地図をもとにランドマークの位置を予測するものが多かった。これは、点や線などの単純なランドマークは識別性が低く、予測なしでセンサデータとランドマークを対応づけるのが難しいためである。しかし、地図が不正確だったり環境が変化したりした場合、この予測は外れる可能性が高い。

これに対処するため、友納らは、識別性の高いランドマークとして家具などの物体を用いて、物体認識によりランドマークの対応づけを行った[16]。そこでは、主要な物体を基準にして、その局所座標系でロボットの自己位置を表す。ロボットは、物体を認識するたびに、その物体からの相対姿勢に基づいてパーティクルフィルタを計算し、自己位置推定を行う。

図3にナビゲーションの様子を示す。なお、このシステムでは、3章の方法とは違い、人手で作成した物体モデルを用いてエッジ線分の対応により物体認識を行っている。

5. 物体モデルを用いた把持動作

従来、マニピュレータによる物体把持では、事前に作成した物体モデルに対して把持方法を定義しておき、センサデータと物体モデルを照合して物体把持を行うものが多かった。これに対し、山崎らは、自動生成した物体モデルに対する把持計画手法を提案した[17]。この方法を用いれば、センサデータから生成した物体モデルを用いて、把持箇所や手先姿勢を自動的に定義できる。また、初見の物体をその場でモデル化し、その把持計画を立てることも可能になる。

図 3 物体認識によるナビゲーション。上段：走行中のロボット，下段：ロボットのカメラで撮影した物体の認識結果

図 4 把持計画の例。上段：シミュレーション結果。下段：実画像から生成したモデルによる把持結果

この方法では、まず、物体モデルを方向付き点群で表す。方向付き点群は、物体形状を構成する密な三次元点に、物体表面に対する法線成分を持たせたものである。さらに、方向付き点群をボクセル空間に投票し、ボクセル集団として再構成することで、データ量の削減と誤差の平均化を図る。

この物体モデルでは、三次元点に法線方向が付加されているため、グリッパと物体モデルの接触状態が評価しやすくなる。マニピュレータの把持姿勢は、グリッパと物体モデルの接触面積、把持位置における物体のバランス、および、手先姿勢の誤差の修正しやすさ [18] を評価して求める。

図 4(a)～(c) にシミュレーション結果を示す。図 4(a) のように、凹みのない直方体では、中央部分を上から把持する姿勢が得られた。一方、図 4(b) のように、中央に凹みのある直方体では、接触面積を大きくするために凹みを避け、かつ重力バランスをとるような姿勢が得られた。図 4(d), (e) は、実画像から生成したモデルで把持を行った例である。図 4(d) は、物体とロボットの距離が近い場合であり、コップの取っ手を避けて上方向から把持している。図 4(e) は、物体とロボットの距離が遠い場合であり、横方向からの把持姿勢を選択している。

6. 物体モデルを用いた可動物の操作

多くの家具は引き出しや引き戸などの可動部をもつが、前

図 5 引き出しの操作の例。左列：操作方法の教示。中央列：前板モデリングの様子。右列：ロボットによる収納操作

述の物体モデル生成では可動部までは復元できない。これに対して、山崎らは、物体形状に可動部の操作情報を加えた IM モデル (Instructed Motion Model) を用いて、可動部をもつ物体の操作を行う方法を提案した [19]。

IM モデルは、次のように生成する。まず、引き出し、引き戸、開き戸などの操作時の手先の動きを分類して、ロボットに一般的な操作知識として与えておく。次に、個々の家具について、人間がロボットの手先を直接動かして、可動部の操作方法を教示する。このとき、ロボットは、教示された操作軌跡データから前述の一般化操作知識に対応する部分を抽出し、可動部の操作モデルとする。さらに、ロボットは、教示中にカメラとレーザスキャナを用いて、可動部の三次元モデルを生成する。このように、IM モデルは、人間が直接教示することによってロボットに与えられるので、特別な道具を用いたり、複雑な操作手順を記述したりする必要がないという利点がある。

図 5 に実験の様子を示す。図 5(A1)～(A3) は、人間がマニピュレータを手に取り、操作方法を教示する様子である。図 5(B1)～(B3) は、引き出しの前板のモデリングの様子である。まず、図 5(B1) で抽出した画像エッジから図 5(B2) に示す前板の候補を得た。そして、操作後の引き出しの位置の予測結果と画像エッジとの対応をとることで、図 5(B3) に示すような操作後の前板候補を得た。図 5(C1)～(C3) は、IM モデルによる収納操作の様子である。IM モデルに基づいて、ロボットは収納作業の開始位置まで移動し、キャビネットの引き出しを開けた後、台車上のペットボトルをキャビネットに入れることに成功した。

7. 物体モデルを用いた地図構築

従来、移動ロボットの地図構築は環境の全体形状をひとかたまりに表すものが多かった。しかし、物体操作を伴う作業を自律的に行うには、物体単位に分割された地図が望

図6 物体地図の例

ましい。友納は、ロボットが走行しながら撮影した画像から認識した物体のモデルを床面に配置することで、物体地図を構築する手法を提案した[20]。

未知環境をカメラだけで走行するのは現状では難しいため、レーザスキャナによる二次元SLAMを併用する[21]。まず、スキャンマッチングにより二次元地図を構築する。スキャンマッチングは、隣接するレーザスキャンを同一領域が重なるようにつなぎ合わせていく技術である。そして、得られた二次元地図を用いて、物体が存在しそうな箇所の画像を撮影し、その画像から認識した物体モデルを床面に並べていく。図6に物体地図の構築例を示す。この地図は、2種類の机、流し台、冷蔵庫、移動ロボット、小机から構成されている。

8. ま と め

物体モデリング・認識の手法を簡単に説明し、ナビゲーションや物体操作などへの応用を紹介した。ロボットが自律的に行動するために、物体認識は本質的に重要である。近年、物体モデリング・認識技術は大きく進歩し、ロボティクスでの研究も盛んになりつつある。しかし、人間がもつ認識能力はすばらしく、ロボットはまだ遠く及ばない。実用面での課題として処理効率とロバスト性のさらなる向上があるが、長期的にはカテゴリー認識や物体概念の獲得などが重要な課題であろう。

参 考 文 献

- [1] Ch. Borst, M. Fischer and G. Hirzinger: "A Fast and Robust Grasp Planner for Arbitrary 3D Objects," Proc. of ICRA1999, pp.1890-1896, 1999.
- [2] A. Hauck and N. O. Stöfeler: "A Hierarchic World Model Supporting Video-Based Localization, Exploration and Object Identification," Proc. of ACCV'95, pp.III-176-III-180, Dec. 1995.
- [3] A. Kosaka and J. Pan: "Purdue Experiments in Model-Based Vision for Hallway Navigation," Proc. of Workshop on Vision for Robots in IROS'95, pp.87-96, 1995.
- [4] L. Petersson, P. Jensfelt, D. Tell, M. Strandberg, D. Kragic and H.I. Christensen: "Systems Integration for Real-World Manipulation Tasks," Proc. of ICRA2002, pp.2500-2505, 2002.
- [5] T. Tsubouchi and S. Yuta: "Map Assisted Vision System of Mobile Robots for Reckoning in a Building Environment," Proc. of ICRA'87, pp.1978-1984, 1987.
- [6] J. Shi and C. Tomasi: "Good Features to Track," Proc. of CVPR'94, pp.593-600, 1994.
- [7] R. Hartley: "In defense of the eight-point algorithm," IEEE Trans. PAMI, vol.19, no.6, pp.580-593, 1997.
- [8] C. Tomasi and T. Kanade: "Shape and Motion from Image Streams under Orthography: A Factorization Approach," Int. J. of Computer Vision, vol.9, no.2, pp.137-154, 1992.
- [9] A.W. Fitzgibbon and A. Zisserman: "Automatic Camera Recovery for Closed or Open Image Sequences," Proc. of ECCV'98, 1998.
- [10] 山崎, 友納, 坪内, 油田: "移動ロボット搭載のカメラを用いた物体の三次元モデリング", 日本ロボット学会誌, vol.23, no.7, pp.868-877, 2005.
- [11] M. Fischler and R. Bolles: "Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography", Communications ACM, vol.24 pp.381-395, 1981.
- [12] K. Kutulakos and S. Seitz: "A theory of shape by space carving," Proc. of ICCV'99, pp.307-314, 1999.
- [13] M. Tomono: "Dense Object Modeling for 3-D Map Building Using Segment-based Surface Interpolation," Proc. of ICRA2006, pp.2609-2614, 2006.
- [14] D.G. Lowe: "Object Recognition from Local Scale-Invariant Features," Proc. of ICCV, 1999.
- [15] 友納: "2次元エッジ特徴と3次元エッジ形状の統合による物体認識", 日本ロボット学会誌, vol.25, no.6, pp.997-1004, 2007.
- [16] 友納, 油田: "不正確さを許すマップと単眼ビジョンによる物体認識に基づく移動ロボットの屋内ナビゲーション", 日本ロボット学会誌, vol.22, no.1, pp.83-92, 2004.
- [17] K. Yamazaki, M. Tomono, T. Tsubouchi and S. Yuta: "A Grasp Planning for Picking up an Unknown Object for a Mobile Manipulator," Proc. of ICRA2006, pp.2143-2149, 2006.
- [18] K. Yamazaki, M. Tomono, T. Tsubouchi and S. Yuta: "Motion Planning for a Mobile Manipulator Based on Joint Motions for Error Recovery," Proc. of IROS2006, pp.7-12, 2006.
- [19] K. Yamazaki, T. Tsubouchi and M. Tomono: "Modeling and Motion Planning for Handling Furniture by a Mobile Manipulator," Proc. of IROS2007, pp.1926-1931, 2007.
- [20] M. Tomono: "3-D Object Map Building Using Dense Object Models with SIFT-based Recognition Features," Proc. of IROS2006, pp.1885-1890, 2006.
- [21] 友納: "ユークリッド変換に不変な特徴量を用いた二次元大域スキャンマッチング方式", 日本ロボット学会誌, vol.25, no.3, pp.390-401, 2007.

友納正裕 (Masahiro Tomono)

1985年東京大学大学院修士課程修了。1985～1999年NEC。2002年筑波大学大学院博士課程修了。2002～2006年科学技術振興機構さがけ研究者。2006～2008年東洋大学。2008年千葉工業大学未来ロボット技術研究センター。情報処理学会, IEEEの会員。

(日本ロボット学会正会員)

山崎公俊 (Kimitoshi Yamazaki)

2007年筑波大学大学院博士課程修了。2006年日本学術振興会特別研究員。2007年東京大学大学院情報理工学系研究科特任助教。2008年同大学IRT研究機構特任助教。情報処理学会, IEEEの会員。(日本ロボット学会正会員)