

### 3次元VTNモデルとその面関係を用いた 物体の形状・姿勢推定手法の提案

岩田 和也<sup>†</sup> 早瀬 光浩<sup>†</sup> 青木 公也<sup>‡</sup>

中京大学 情報科学研究科<sup>†</sup> 中京大学 情報理工学部<sup>‡</sup>

#### 1 序論

本稿では、3次元形状モデルとその面関係を用いた物体の形状と姿勢の推定手法を提案する。提案手法で用いる可変テンプレートネットワークモデル(Variable Template Network Model: 以下、VTNモデルという)[1]とは図形的推論の一種で、長方形や楕円などの基本形状の図形に拡大・縮小や回転といった運動を作用させたものである。これを拡張した3次元形状の2次元投影像を用いたVTNモデルによる物体認識には成功例[2]があり、野菜や飲料缶などの認識に成功している。

ロボットビジョンにおける物体認識には、精度と処理速度が求められるが、従来手法では2次元投影像のみを用いてモデルとのマッチングを行っていたため、精度に欠ける面があった。そこで提案手法では直方体のモデルに、それを構成する各面の情報と面同士の関係を持たせ認識に用いる。これによって精度と処理速度の向上を目的としている。

本稿では、ロボットが身近にある物体を把持するというタスクを想定し、物体認識を行う手法について述べる。

#### 2 VTNモデルを用いた認識手法

提案手法の認識プロセスはトップダウン処理とボトムアップ処理、それらの結果を用いるマッチング処理からなる。Fig.1に処理の流れを示す。

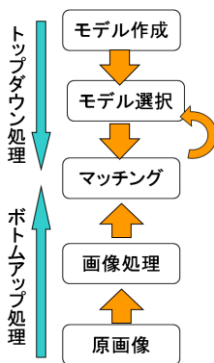


Fig.1 処理の流れ

ボトムアップ処理とは、入力画像に対し画像処理を施し特徴量と抽出することであり、トップダウン処理とは、VTNモデルを作成することである。

#### 2.1 トップダウン処理

トップダウン処理では、VTNモデルの作成と運動の定義を行う。

VTNモデルは基本形状となる図形とその運動列からなり、基本形状に運動を作用させて構成される。その形状については、点、線分、面などに抽象化され、離散化されている。運動については、拡大、縮小、回転などに抽象化され、離散化されている。形状と運動列をネットワークで構成することにより、運動を簡単に表現でき、形状を柔軟に変化させることができる。推論はこのネットワークをたどることによって表現させる。これにより、形状と運動を同じように表現できる。Fig.2に、VTNモデルのネットワークの概念を示す。

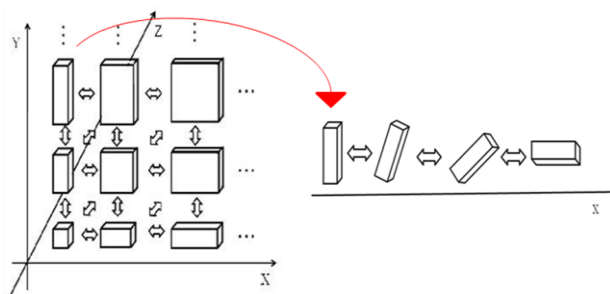


Fig.2 VTNモデルのネットワークの概念

提案手法で用いたVTNモデルは、直方体を基本形状とした。5×5×5voxelを最小サイズとし、X軸、Y軸、Z軸それぞれに5voxelずつ10段階の拡大縮小の運動を持つ。各モデルはX軸に関して0°~90°、Y軸に関して0°~80°の範囲で10°ずつ回転運動を持つ。これらを2次元投影し、各運動をリスト構造で接続することで表現している。各モデルはサイズや回転角度の他に、見えている面数、各面の面積、重心間の距離や位置関係といった面に関する情報をパラメータとして持つ。モデル総数は90000個となる。

#### 2.2 ボトムアップ処理

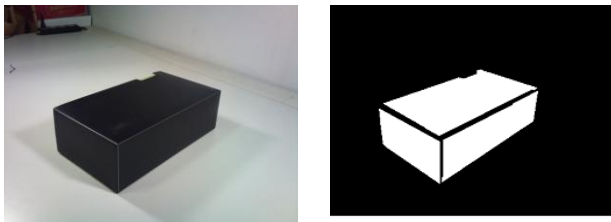
ラスタ形式で表現された入力画像に対し、形状と特徴量で表現されたモデルをそのまま当てはめ、マッチングを行うことは不可能である。入力画像には、ノイズなどの認識の障害となるものが存在しているためそれらを除去し、マッチングに必要な特徴量を得るためにボトムアップ処理を施す。本手法では、2値化、最大成分抽出、

Presumption of Shape and Posture Using Variable Template Network Model and Their Relation of surface

<sup>†</sup>Kazuya Iwata, Mitsuhiro Hayase  
Graduate School of Computer and Cognitive Sciences,  
Chukyo-University

<sup>‡</sup>Kimiya Aoki  
School of Information Science and Technology,  
Chukyo-University

エッジ抽出, 線分検出, 面抽出を行った. 入力画像の例と, ボトムアップ処理の結果を Fig.3 に示す.



(a)入力画像 (b)処理結果  
Fig.3 ボトムアップ処理の結果

ボトムアップ処理の結果から, モデルが持つパラメータと同様の特徴量を得る.

### 2.3 マッチング

トップダウン処理によって作成した VTN モデルと, ボトムアップ処理を施した処理画像とのマッチングを行う. 前述したようにモデルの総数は 90000 個であるため, すべてのモデルに対して重なり率を計算し, 最適モデルの選択を行っている膨大な時間がかかってしまう. そのため, 本手法では見えている面の数ごとに場合分けしてマッチングを行う. 面数によってある程度の回転角度を推定することができるためである. 見えている面が 1 面, 2 面, 3 面の各場合でモデルの探索を行う範囲と用いる特徴量が異なる. これによって無駄な探索を省き, 効率のよいマッチングを行うことができる.

マッチングはモデル群のネットワークを辿り, 最適となるモデルを探索する. マッチングの際には, VTN モデルのパラメータと処理画像の特徴量を比較し, 面数の値と各面の位置関係が一致し, 各面の重心間の距離が一定の割合以上一致した場合のみ, モデルの 2 次元投影像と画像中の対象領域の重心を一致させたときの重なり率の計算を行う. 概念図を Fig.4 に示す.

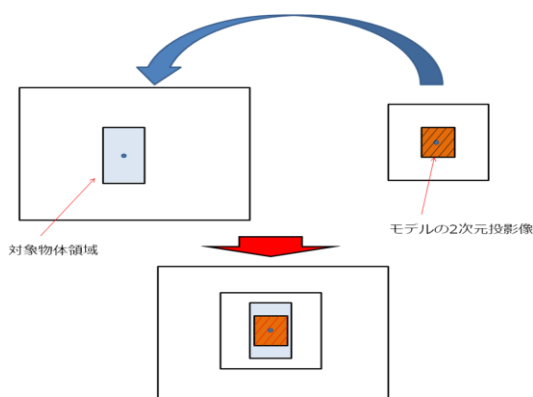


Fig.4 マッチング処理

重なり率は, 対象となる領域面積とモデルの面積に対し, 式 1 で定義される.

$$\text{重なり率} = \frac{\text{重なり面積}}{\text{モデル面積}} \times 100(\%) \quad (\text{式 1})$$

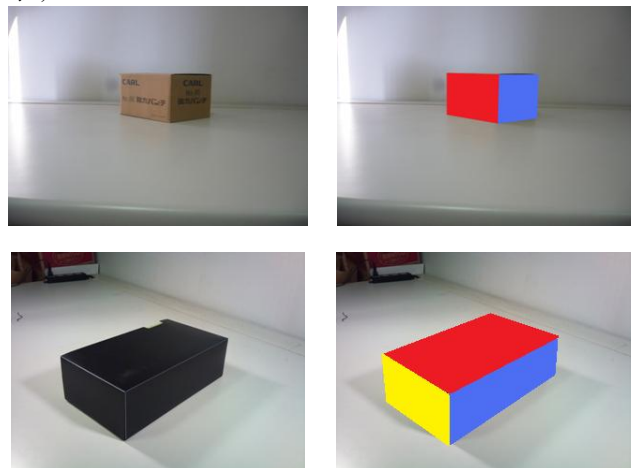
面数ごとに探索に用いるしきい値が異なり, 1 面, 2

面, 3 面のそれぞれの場合で 90%, 85%, 80%以上で, 重なり面積が最大となるモデルを最適モデルとして選択し, 探索を終了する.

## 3 実験と考察

### 3.1 静止画を用いた実験

本研究では前述した通り, ロボットが物体を把持するというタスクことを想定している. そのため, 一般的な室内の照明条件下で机にある長方形の物体を撮影した静止画を入力画像として実験を行った. 結果を Fig.4 に示す.



(a)入力画像 (b)マッチ結果  
Fig.5 実験結果

### 3.2 評価

各面の見え方ごとに 30 シーンずつ実験を行い, 得られた結果の精度, 処理時間の平均を表 1 に示す.

表 1 実験結果

面積一致率	94.2%
角度一致率	X 軸 : 87.3%, Y 軸 : 86.1%
マッチング時間	19.3ms

これらの値は, 2 次元投影像のみを用いた手法と比較して, 面積の精度は同程度となったが, 回転角度の精度に関しては 20%以上, マッチング時間に関しては 1/15 と大幅な向上を達成することができた.

## 4 結論

本稿では VTN モデルとその面関係を用いることで, 高精度な 3 次元物体の認識が可能であることを示した. 今後は他の 3 次元形状へと応用し, それらの組み合わせによって複雑な形状の物体も可能になると考えられる.

- [1] 嶋田晋: "時空間モデルを用いた図形的推論体系の提案", 人工知能学会全国大会論文集, Vol.9, pp.41-44, 1995.
- [2] Mitsuhiro Hayase and Susumu Shimada: "Posture Estimation of a Human Body from Thermal Images of 2D Appearance Models of 3D Ellipsoidal Model", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.13 No.3, May 2009.