

3 次元 VTN モデルとその組み合わせによる物体認識

早瀬 光浩 †

† 中京大学大学院情報科学研究科

嶋田 晋 ‡

‡ 中京大学情報理工学部

1 はじめに

近年、様々な分野にロボットが進出してきている。ロボットが人間と共存するためには、物体を把持・移動することが必要である。物体を把持・移動するためには、物体の認識、位置・姿勢の認識は重要なタスクである。物体認識は CV (Computer Vision) や RV (Robot Vision) の分野で最も重要な課題の一つである。特に実世界シーンの画像に対して、ロボットがその中に含まれる物体を一般的な名称で認識することや位置・姿勢を認識することは基本的な課題である。物体認識の手法は大きく分けて二つの方法がある。

一つめは、画像の特徴量を利用した Bag-of-Feature や Context-based があげられる。これらは、事前に学習したデータをもとに画像中にどのような物体があるのかを分類している。この手法は、多くの成果をあげ、ある程度の物体を認識することができている。しかし、物体がどのような位置であるのかを認識するのは困難である。

二つめは、モデルベースドビジョンがある。これは、あらかじめ物体のモデルを CAD などのデータで作成しマッチングを行う方法や固有空間法などのように、物体ごとのアスペクトのモデルを用いて認識する方法があげられる。このモデルベースドビジョンの方法は、特定物体の認識では成果をあげているが、認識したい物体のモデルをあらかじめ持っている必要があり、また、世の中にあるものをすべて認識しようとすると膨大なモデルが必要になる。

本研究は、このモデルベースドビジョンのアプローチに基づき、3 次元 VTN (Variable Template Network) モデルとその組み合わせによる物体認識を行っている。我々の研究室では、基本形状に拡大・縮小・回転等の変化を運動として離散的に作用させた VTN モデルを提唱している。2 次元 (2D) の基本形状とその組み合わせによる研究として、岩田らによる手指の認識がある [1]。これは位置関係を考慮し、手指の認識を行っている。3 次元 (3D) VTN モデルによる研究として、早瀬らによる楕円体モデルを用いた人体の胴体部の認識 [2]、円

柱モデルを用いた飲料缶の認識等がある。これは、3D 基本形状モデルに対して、拡大・縮小を与え、複数の比率の 3D モデルの見え方モデルを得る。得られた見え方モデルをネットワークで接続したモデルを 3D VTN モデルとしている。これらの研究は、単一形状の認識ではある程度の成果が得られている。

本稿では、3D VTN モデルとその組み合わせによる物体認識の手法について述べる。取得した画像に対して領域分割を行い、処理した画像と 3D VTN モデルとのマッチングを行う。複雑な形状の物体でも、基本形状に分割しその組み合わせで認識できる。

2 章では、VTN モデルの構成方法を述べ、3 章では領域分割と VTN モデルを用いた認識手法について述べる。4 章では、実験結果ならびに評価、そして、5 章でまとめを述べる。

2 3 次元 VTN モデルの構成

2.1 3 次元基本形状モデルの構成

初めに、基本となる 3D 基本形状モデルを作成する。3D モデルは、OpenGL を用いて作成を行った。画角は、使用するカメラと同じ画角に設定した。今回作成した基本形状は、楕円体・円柱・直方体の作成を行った。基本形状は、OpenGL の論理単位で半径 0.5、高さ 1 で作成した。そして、 $5 \times 5 \times 5$ の大きさを最小の 3D 基本形状モデルとし、モデル座標系の X 軸、Y 軸、Z 軸に沿って、それぞれ 5 ずつ拡大を行い、最大 $25 \times 25 \times 25$ の大きさの 3D 基本形状モデルを作成した。拡大・縮小された 3D 基本形状モデルは、合計 125 モデル存在する。概念図を図 1(a) に示す。

2.2 2 次元見え方モデルの構成

次に 2.1 で述べた 3D 基本形状モデルの一つ一つに対して、様々な角度から 2D 見え方モデルを得る。見え方モデルは、3D 基本形状モデルに対して、視点は固定し、モデル座標系の X 軸・Z 軸に対して 0 度から 10 度ずつ 180 度まで回転を行い、投影する。これにより一つの 3D 基本形状モデルから $324(18 \times 18)$ モデルの 2D 見え方モデルが得られる図 1(b)。2D 見え方モデルは、すべての 3D 基本形状モデルから得るので、合計 $36,125(125 \times 324)$ の 2D 見え方モデルとなる。

Object Recognition Based on 3D VTN Model and these Combination

†Mitsuhiro HAYASE ‡Susumu Shimada

†Graduate School of Computer and Cognitive Science, Chukyo University

‡School of Information Science and Technology, Chukyo University

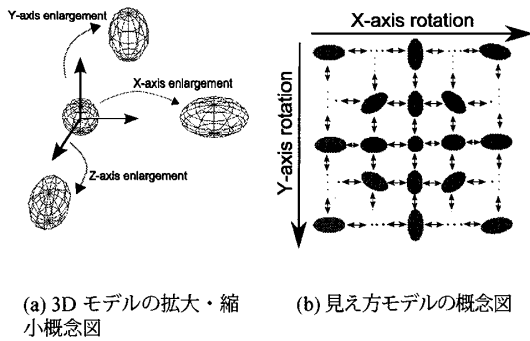


図 1: VTN モデルの構成

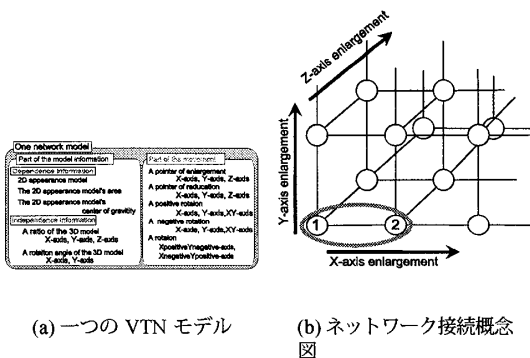


図 2: ネットワークの構成

2.3 ネットワークの構成

これら 2D 見え方モデルが独立に存在するのではなく、ネットワークで接続されネットワークモデルとして存在する。ネットワークの一つのモデルに注目すると、図 2(a) に示すようにモデル情報部と運動部からなる。そして、モデル情報部は、形状に依存する部分と依存しない部分の情報からなる。依存する部分は、2D 見え方モデル、そのモデルの面積と重心を保持する。依存しない部分は、見え方モデルを得たときの 3D 基本形状モデルの大きさや回転角度のパラメータを保持する。運動部は、3D 基本形状モデルの大きさや回転角度が変化したときの 2D 見え方モデルへのポイントを保持する。ネットワークでつながれたモデルは、14 方向へのポイントを持つ。ここでいう方向とは、3D 基本形状モデルのパラメータが変化したときに得た 2D 見え方モデルへのポイントのことである。図 2(b) にネットワーク構造の概念図を示す。図中のノード部分にそれぞれ様々な角度から得られた 2D 見え方モデルを持っている。例として図 2(b) 中の 1, 2 と書かれた部分では同じ回転角度の部分と接続している。

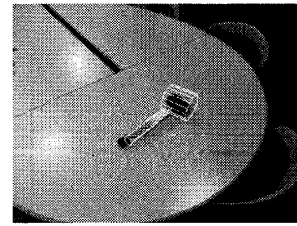


図 3: 実験結果。

3 提案手法

本システムは、トップダウン処理とボトムアップ処理の二つからなる。トップダウン処理は、モデル作成とモデル探索からなる。モデル作成は、3D 基本形状モデルに拡大・縮小を行い、様々な角度から 2D 見え方モデルを得る。それらの見え方モデルをネットワークで接続する。ボトムアップ処理は、入力画像に対して、領域分割を行う。領域分割は、隣接する画素の HSV 空間上で対応する 2 点間の距離を求めしきい値以上のものをエッジとし、ラベリング処理を行い領域分割画像を得る。領域分割画像と 2D 見え方モデルとのマッチングを行い、最適な 2D 見え方モデルを探索する。最適な 2D 見え方モデルのパラメータから 3D 基本形状モデルを作成し表示する。マッチングは、各領域ごとに繰り返す。

4 実験と考察

本研究では、入力画像を 320×240 、2D 見え方モデルを 120×120 とした。実験結果を図 3 に示す。実験結果より、各パーツごとにほぼ正しい位置でマッチしていることが分かる。

5 まとめ

本稿では、3次元 VTN モデルとその組み合わせによる物体認識の手法について述べた。領域分割した画像に対して、VTN モデルを用いてマッチすることができ、各パーツごとに認識することができた。今後は、マッチした各モデルから物体ごとの接続関係を取り出し、物体の名称認識を行いたい。

参考文献

- [1] 岩田, 他, 第 8 回情報科学技術フォーラム講演論文集, H-032, 2009.
- [2] Mitsuhiro Hayase and Susumu Shimada, JACIII, Vol.13, No.3 pp.172-177, 2009.