

VTNモデルとその接続関係を用いた手指の認識 Recognition of Hand and Fingers by 2D Variable Template Network Models and Their Connection Relationship

岩田 和也[†] 早瀬 光浩[†] 嶋田 晋[‡]
Kazuya Iwata Mitsuhiro Hayase Susumu Shimada

1. はじめに

近年、ロボットは様々な場所に進出してきている。人間とロボットが共存していくうえで、コミュニケーションを図る場合、様々な方法が考えられる。人間側が直感的にわかりやすい方法として手で指示を与えることが挙げられる。そのためには、人の手のような複雑な形状の物体をロボットが認識する必要がある。

手認識の研究は数多く行われており、トップダウン的に手のモデルをもつ研究[1]や、ボトムアップ的に画像中の特徴量を抽出し手を認識する研究[2]などがある。

また複雑形状を認識する類似研究として伊藤による幾何学的関係モデルと意味表現情報を用いた認識の研究例[3]がある。この研究では動物のような姿勢の変化により形状が変化する物体も、基本形状に分割し、関連付けることにより認識が可能になることが示されている。

また接続関係を用いている研究例として、影山らによるパーツ分解とカテゴリ分類の研究例がある[4]。これはシルエット図形から中心軸をもとに分解・再構成を行い、最大面積を持つパーツを主パーツとして定義することにより、その他のパーツがどこに配置しているかをという情報を基に複雑形状物体の認識を行っている。

筆者らの研究室では、図形推論の一環として「可変テンプレートネットワークモデル(以下、VTNモデルと呼ぶ)」を提唱している[5]。これは2次元の長方形に拡大、縮小、回転の運動を作用させた可変2次元モデルである。さらにこれを拡張した3次元モデルの見え方モデルを用いた研究として早瀬による事例がある[6]。これは楕円モデルを用いて人間の胴体部分の認識を行っている。

そこで、本稿では入力画像に画像処理を施し、処理画像に対してVTNモデルを用いパーツごとにマッチングを行い、それらの接続関係を用いて手の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

本研究で作成したシステムはトップダウン処理とボトムアップ処理からなる。ボトムアップ処理とは入力画像に対しさまざまな画像処理を施し、トップダウン処理とはVTNモデルの運動を定義することである。それぞれの処理の接点としてマッチングを行い、マッチ結果からモデルの接続関係を用いて手の判断を行う。接続関係はあらかじめ知識として与えられ、掌と指の位置関係をもとに入力された画像が手であるのか、手ではないのかの判定をする (Fig.1)。

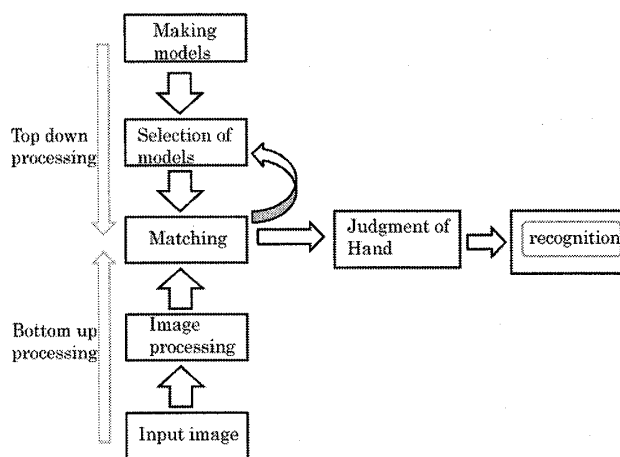


Fig.1 The outline of recognition

2.1 トップダウン処理

2.1.1 VTNモデルによる推論

人間は対象物体を自由に運動させ推論を行うと考えられる。そこで図形による推論の知識表現と推論体系として新たに「VTNモデル」を利用する[5]。

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。VTNモデルによる推論は次の三つのステップによる繰り返して行われる。

Step1: VTNモデルにおいて、ある時点で可能な運動を列挙する

Step2: ある規準(解くべき問題に依存)に従って、その中から運動列を選択する

Step3: 運動列を作用する

具体的認識における推論方法では、Step1における可能な運動列とはモデルの拡大、縮小、回転などである。

[†] 中京大学大学院 情報科学研究科 Graduate School of Computer and Cognitive Sciences, Chukyo-University

[‡] 中京大学 情報理工学部 School of Information Science and Technology, Chukyo-University

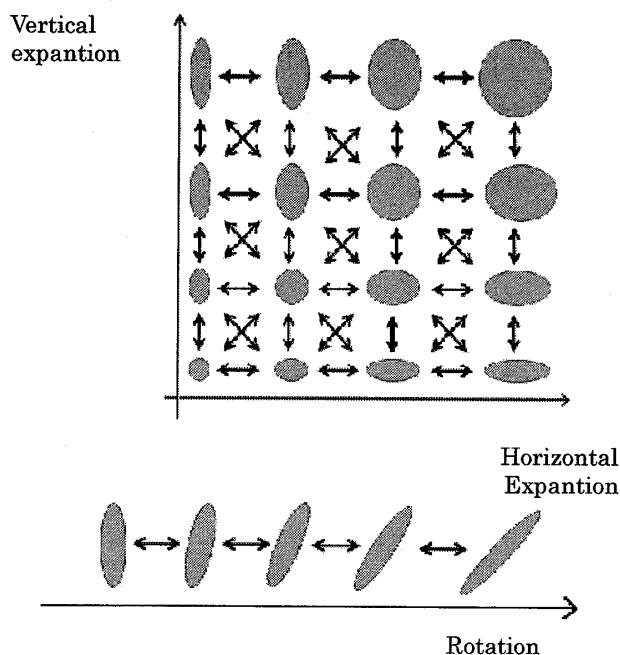


Fig.2 Transformation of models

また、Step2のある基準とは、“図形との重なり率が大きくなる方向”を基準とするのが一般的である。この柔軟に変形できる図形をデータ構造として取り扱い、それをもとにした推論を行う。

このVTNモデルを用いて複雑形状を基本形状に分割し、推論を進めることにより、固定の手のモデルをもつ認識手法と比較して高い汎用性を持つことができた。

2.1.2 VTNモデルの構成

今回使用したVTNモデル(以後、モデルと呼ぶ)は最小単位として半径が4pixelの円を基本形状とし、縦横それぞれ10段階に拡大モデルを持ち、それぞれのモデルが縦・横・斜めの拡大・縮小に対応するリスト構造で連結されている(Fig.2)。また、各々は左右10度ずつ、180度回転し、上下の区別は行わない。また、各モデルはパラメータとしてモデルの面積、重心、角度の情報が与えられている。

2.2 ボトムアップ処理

入力された画像にはノイズなどの不必要な情報が含まれているため、ボトムアップ処理として各種の画像処理を施し、認識すべき対象である手の領域を抽出する。具体的には、人間の肌の色を抽出するのに適したIRgBy色抽出、2値化、エッジ抽出を行い、収縮・拡散を繰り返すオープニング処理によって手から掌部分を抽出する。これと元の画像の差分をとることにより指部分を抽出し(Fig.3)それぞれにラベリング処理を行う。その後重心抽出、主軸抽出などの特徴量抽出を行う。(Fig.4)

2.3 マッチング

ボトムアップ処理を施した入力画像に対して、トップダ

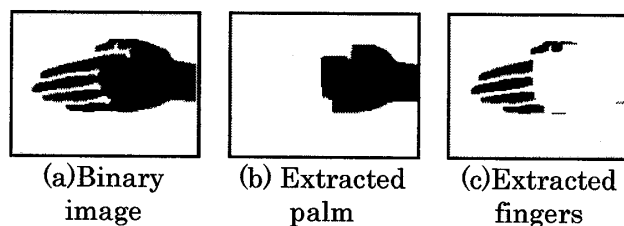


Fig.3 Separate fingers from hand

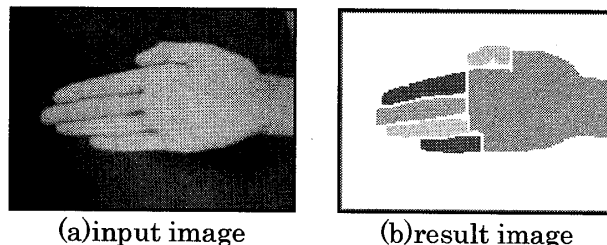


Fig.4 Result of image processing

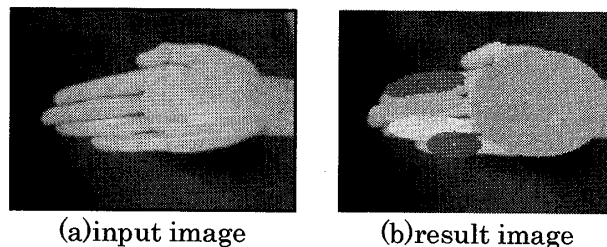


Fig.5 Result of matching

ウン処理によって生成したモデルとのマッチングを行う。本研究ではマッチングの手法として面積マッチングを用いている。これは図形とモデルとの重なった面積を基準とする手法である。図形とモデルの重なり率を基準として推論を進めていく。

ボトムアップ処理により得られた対象図形の重心と慣性主軸を、モデルの重心と回転角度と重ね合わせ、一定の重なり率以上で最大の面積となるモデルを最適マッチモデルとし、掌部分とすべての指部分にマッチするように繰り返す(Fig.5)。

2.4 接続関係を用いた手判断

2.3で得られたマッチ結果から、マッチした物体が手であるかどうかを判断する。掌部分にマッチしたモデルの慣性主軸に対して、垂直に交わる直線を基準線とする。この直線に対して指部分にマッチしたモデルの重心がすべて同じ側にある場合に手であると判断し、一つでも違う側にある場合には手と判断されない(Fig.6)。また、手である場合には掌にマッチしたモデルと指にマッチしたモデルの位置関係から、手がどちらの方向を指しているかを判断する。

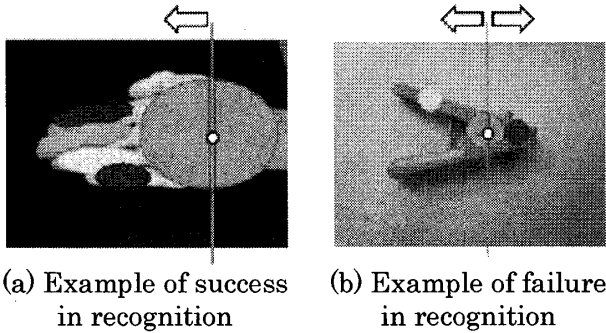


Fig.6 Result of judgment

3. 静止画を用いた実験

3.1 実験の前提

本研究では人間がロボットに手で指示を出す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、フレームサイズは 160×120pixel のカラー画像を入力画像とした。実験に用いたものとして開いた手、閉じた手、ポーズをとった手など、また手以外のものを排除するために洗濯ばさみなども用いている。

3.2 実画像マッチ結果

実際に手の画像、および手ではない物の画像を入力画像 (Fig.7)として用い、マッチングを行った。結果を以下に示す (Fig.8)。

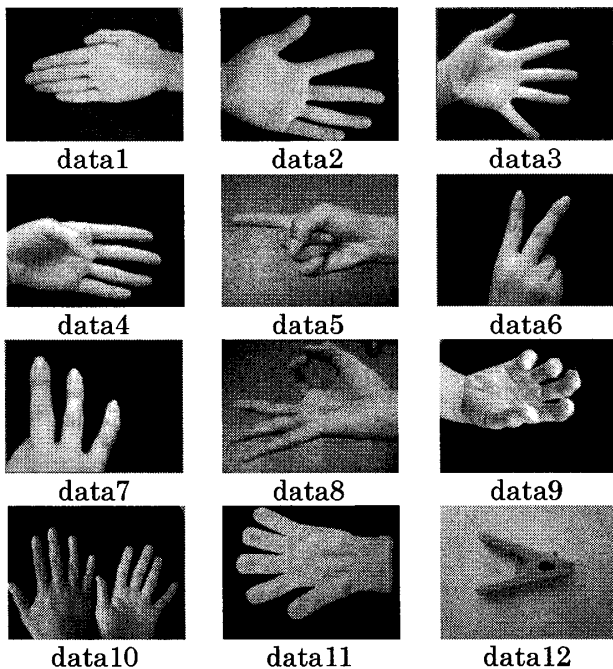


Fig.7 Input image

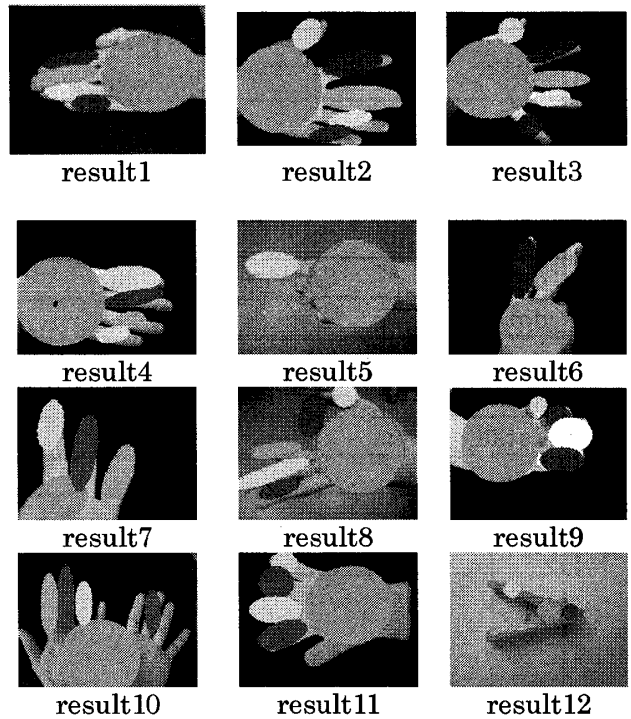


Fig.8 Match result image

data1 から data4 は開いた手, data5 から data9 は手でポーズをとったもの, data10 は 2 つの手である。また今回 data11, data12 の様な手ではないものの画像についても実験を行った。これらは本来、IRgBy 色抽出の肌色抽出では抽出されないため、マッチングを行うことは不可能だが、色抽出のパラメータの範囲を意図的に変更して抽出を行い、マッチングを行った。

3.3 手判定結果

3.2 のマッチ結果から手であるかどうかを判定する。結果を以下に示す (Table.1)。data9 の手の場合、手首や手の平が正面を向いていないため掌部分の主軸角度から手ではないと判断されたと考えられる。また、data10 の 2 つの手の場合には、掌抽出の段階で 2 つの掌が抽出されたために上手くマッチせず、結果として手ではないとされてしまったと考えられる。

3.4 評価・考察

3.3 に示した通り、手判定の正解率は 80%となった。data9 のように掌や指がカメラに対して垂直に向いている場合、先行研究としてある 3次元の見え方モデルを用い、また data10 のような 2 つの手の場合の誤認識には、また掌抽出の方法を改め、各々の手のひらに対応した接続関係を与えることにより解決すると考えられる。

処理時間としてはマッチ時間が平均 8.72ms であり、すべてのモデルを総当たりで探索を行った場合には約 1200ms であったため、非常に高速であるといえる。これによって VTN モデルの有効性を示すことができ、よってリアルタイム処理が可能になると考えられる。

Table.1 Result of judgment

image	Judgment result	Right or wrong
data1	○	○
data2	○	○
data3	○	○
data4	○	○
data5	○	○
data6	○	○
data7	○	○
data8	○	○
data9	×	×
data10	×	×
data11	○	○
data12	×	○

4 まとめ

本研究では VTN モデルとその接続関係を用いて、掌と指の各部位が認識可能であることを示した。今後は 3 次元見え方モデルへの拡張や、リアルタイム処理を取り入れ、ロボットに搭載し、より実用的な認識が可能となると考えられる。

参考文献

- [1] 島田伸敬, 白井良明, "手指ジェスチャの画像計測手法とその応用", 情報処理学会論文誌 Vol.1 No.1, pp.53-63, 2008.
- [2] 河野武, 伴好弘, 上原邦昭: "画像解析型手認識による片手入力インタフェース" FIT 2003 講演論文集, pp.99-100.
- [3] 伊藤浩士: "複雑形状の 2 次元基本形状への分割と関係づけを用いた対象物体の認識", 中京大学大学院情報科学研究科修士論文, 2004.
- [4] 影山勝彦, 大町真一郎, 阿曾弘具: "シルエット図形のパーツ分解とカテゴリー分類法" 電子情報通信学会論文誌, Vol.J91-D, No.5, pp.1305-1313, 2008.
- [5] 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.
- [6] Mitsuhiro Hayase and Susumu Shimada: "Posture Estimation of a Human Body form Thermal Images of 2D Appearance Models of 3D Ellipsoidal Model", Journal of Advanced Computational Intelligence and Intelligent Infomatics, Vol.13 No3, May 2009.