シルエットを利用した手指の三次元形状推定法†

亀田 能成‡　美濃 導彦§　池田 克夫¶

京都大学工学部‖

手指を対象とし、そのシルエットからモデルの三次元形状の推定を行う方法を提案する。手指は部品位置決定木という構造記述でモデリングされる。手指の形状を決定する複数のパラメータは、部品位置決定木を利用した簡潔なアルゴリズムで求められる。本手法の特徴は、モデルの形状推定時に画像情報として面積を用いるだけで、画像での複雑な計算を行わない点である。実際に実験を行い、シルエットから三次元形状の推定が可能なことを示した。

1　はじめに

対象が既知である三次元物体の姿勢・形状推定において、モデルベースドビジョンが有効であるとされている。これは、視覚システムが、モデルの知識を積極的に利用することで、画像情報を効率良く推定に用いることができるからである。

特に、対象物体の三次元形状が変形しないことを前提とし、その物体の姿勢推定を行うシステムを構築する場合、物体モデルの三次元幾何構造をいかに効率良く利用するかに注意が払われてきた。代表的な研究のひとつとしてACRONYM[1]やそれに続く研究が挙げられる。一般に、三次元物体の姿勢推定を行う場合、画像を信号処理して得られる特徴の量は、膨大となる。これを“ボトムアップ的情報”と呼び、この大量の情報を、モデルから得られる知識に如何に効率良く対応させて姿勢推定を行うかが問題となる。久野らは、ボトムアップ的情報を、画像から一次的に得られる局所的特徴と、それらを組み合わせて得られる中間レベル特徴とに分類し、各々の認識戦略における得失について述べている[2]。

これに対して、認識対象に変形を認めるような研究については、まだあまり研究が進められていない。これは、画像から抽出されるボトムアップ的情報が大量であるのに加え、対応させるべきモデルの知識の増加によって、ボトムアップ的情報のモデルへの対応付けにおける計算量が膨大になってしまうからである。

そこで、本研究では、認識対象に変形を許す代わりに、画像を処理して得られる情報が少ない環境を想定し、そのもとでどのような認識の枠組が必要とされるかを実験を通して考察する。そのための認識対象として、形状にかなりの複雑さをもつ人間の手指を選択した。また、画像情報として、シルエット画像を利用することにした。シルエット画像は、対象が存在する領域では1の画素値をとり、そうでない領域では0をとる。シルエット画像は輝度画像に比べて画像情報量が少ないが、これを用いることにより画像に対する適切な形状推定の範囲が広がる。本手法では、出来るだけボト

---

†A 3-D Shape Presumption Of A Human Hand Using A Silhouette

‡Yoshinari Kameda

§Michihiko Minoh

¶Katsuo Ikeda

‖Faculty of Engineering, Kyoto University

ムアップ的情報の量の増大を抑えるため、画像情報は各画素の位置とその値しか利用しない。さらに探索空間を狭めるため、効率の良いモデリングを行う。実験の結果、このような環境設定下で、三次元形状の推定が可能であることを示した。

以下、2章で手指に対する効率の良いモデリング法について考察し、3章でそのモデルを用いた三次元形状推定方法について述べ、4章で実験結果を示し、考察を行う。

2 手指モデル

手指のように複雑な形状をとり得る物体をモデリングする場合、そのモデリングには以下のような性質が望まれる。

- 簡潔なデータ構造である。すなわち、対象の特性によく一致し、その表現に一意性がある。
- 三次元形状推定法のアルゴリズムが、効率良く利用できる。

本節では、人間の手指を解剖学見地から検討し、それに基づいたモデリングを提案する。

2.1 手指の要素分割

本研究で対象とする手指は、人間の前腕中途から遠位の右手指である。この部分は、尺骨、橈骨を含めて29個の骨とそれらの間に存在する関節、及び付随する多数の筋、腱から構成されている。手首も含めて、一般に人間は剛体ではない。これは、人間が多関節を持ち、なおかつ変形する筋肉を有するからである。しかし、筋肉による変形はそれほど大きな変化量を持たないので、関節毎に人体を分割することを考えると、各部品は剛体であると見なせる。このように、手指は幾つかの剛体部品を組み合わせてモデル化することが可能である。

解剖学的な見地からみると、手首を構成する八つの手根骨間に存在する手根間関節及び手根中央関節は各々が殆んど動かないので、この部位での運動は前腕と手指間の運動に含めて考えてよい。掌部分に存在する第2、第3、第4、第5指の中手骨間の中手間関節も殆んど機能しない。また、中手骨と手根骨間の中手手根間関節も可動範囲が狭い。これらの事実に基づき、四つの中手骨と八つの手根骨をまとめて一つの剛体部品とする。前腕についても、手首の近傍では橈骨と尺骨はその位置関係が固定であると見なせるので、一つの剛体部品と考えることにする。結局、本研究では右手指のモデルを17個の剛体部品で記述する。

2.2 手指モデルの構造記述

ここで利用する三次元形状モデルでは、その形状を決定づけるのは関節で繋がる剛体部品どうしの位置関係である。人間には伸縮機能を持つ関節がないので、実質的には手指中の関節の回転角がパラメータとなる。これを関節パラメータと呼ぶ。

前節で除去した関節以外に、手指モデルの17個の剛体部品間には16の関節が存在し、各々一ないし二自由度を持つ。これらのパラメータの総数は22である(表1)。各パラメータには、運動学に基づく可動範囲が定められている。これら22個の関節パラメータを全て求めることは、手指の三次元形状の推定を意味する。

表1: 採用した手指の関節

名称	種別	自由度	個数
指節間関節	蝶番	1	9
母指中手指節間関節	顆状	1	1
中手指節間関節	顆状	2	4
母指中手手根間関節	鞍状	2	1
橈骨手根間関節	楕円	2	1

剛体部品間の関係を関節によって全て記述すれば、それが手指モデルの構造記述と

なる。そこで、本研究では、全ての関節が二つの剛体部品間の二項関係であることに着目し、かつ簡潔なデータ構造が要求されていることから、手指モデルの記述法に木構造を採用する。手指モデルのこの構造記述を、部品位置決定木と呼ぶ。図 1にこれを示す。各剛体部品は木状の接続関係を持つ。各ノードは剛体部品を表し、ノード間のアークは、太線が二自由度の関節を表し、細線は一自由度の関節を表している。この部品位置決定木表現において、各ノードは次のような情報を保持する。

- 親ノードへの一本のリンク
- 子ノードへの複数のリンク
- 親ノード剛体部品と当該剛体部品との間の関節の親ノードローカル座標系での位置 (これを接続点と呼ぶ)
- ローカル座標系 (ただし、親ノード剛体部品との関節の接続点を原点とする)
- 親ノード剛体部品との関節パラメータ (一ないし二変数)
- 関節パラメータの可動範囲
- ローカル座標系で表現された剛体部品の幾何的な形状

剛体部品の形状は、 CG で利用されているどのような形式を用いて表現してもよい。これは、本手法の利点の一つといえる。

部品位置決定木は、三次元形状推定法の制御情報をその構造によって表現している。すなわち、この木構造は、各剛体部品の位置が親ノードから子ノードへ向けて順番に決定されていくことを意味する。また、関節パラメータは全て独立である。これにより、部品位置決定木はモデリングに要求される二つの資質を満たしているとわかる。

2.3 本研究における手指モデル実現

本研究では、剛体部品をパッチによるサーフェスモデルで実現している。これをワイヤフレームで表示したのが図 2である。

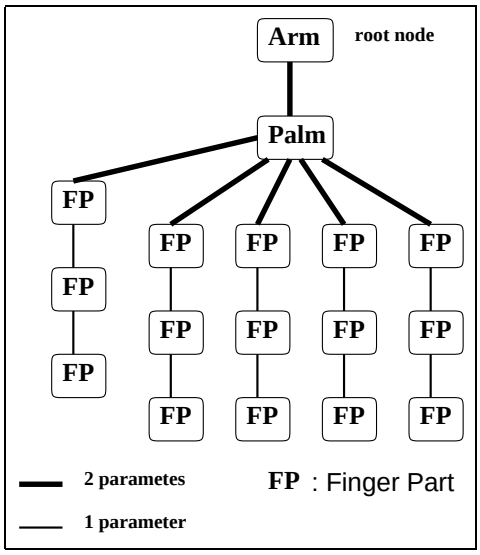

図 1: モデルの部品位置決定木

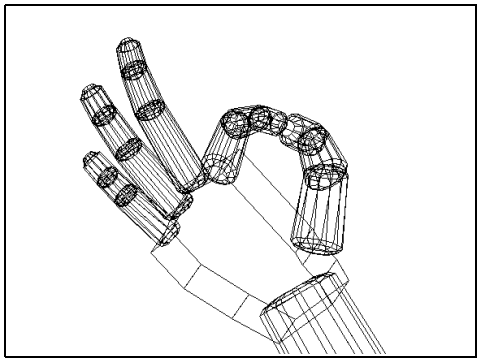

図 2: モデルのワイヤフレーム表示

また、部品位置決定木のルートノードに対応する剛体部品をルート剛体部品と呼ぶ。ここでは、これを中心軸を持つ楕円錐台で表現している。

3 入力画像に対する制約条件

本節では、入力画像に対する制約条件を明示する。本研究で用いる入力画像である手指のシルエット画像は、以下の条件を満たす輝度画像を、背景の最高輝度値を閾値として二値化した画像である。

- 対象は同一人間の右の手指である。手首を含んだ指先までは欠けることなく画像上に撮影される。前腕は中途まで撮影される。
- カメラと対象との距離は一定であり、その距離は既知である。
- 入力画像は、全て同一のカメラパラメータで撮影される。焦点距離に対して手首の奥行きは充分小さい。
- 前腕の中心軸は、画像面と平行である。
- 照明は間接光ないしそれに近い光源を用いる。
- 背景は黒地とし、画像内には手指以外の物体は写されない。
- 画像データは 560×420 ピクセルの解像度を有する輝度画像とし、縦横ともその半分以上の大きさで手指領域が存在する。

図 3に上記の条件で撮影した手指画像を示す。見やすいように輝度値を反転し、コントラストを上げている。これを二値化したシルエット画像が図 4である。

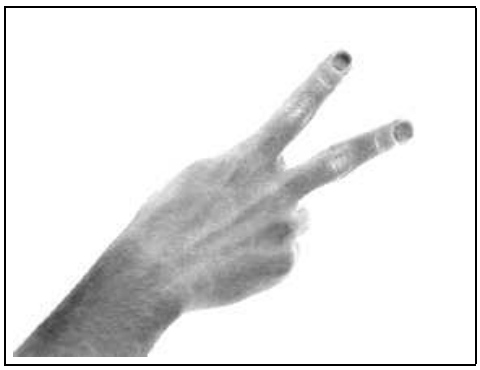

図 3: 原画像

4　手指の三次元形状推定法

本節では、与えられたシルエット画像からルート剛体部品の位置を計算する方法を示し、次いで部品位置決定木を用いて手指の三次元形状を推定する方法について述べる。

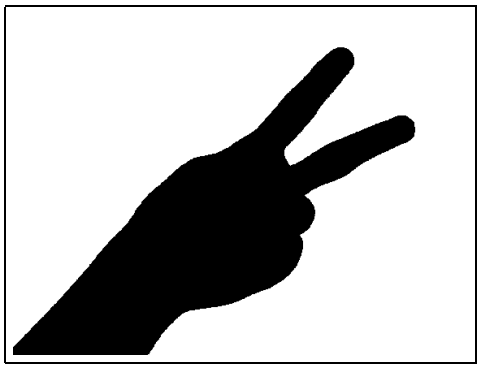

図 4: 入力シルエット画像

4.1　ルート剛体部品の位置決定

ここでは、前節で述べた制約条件を利用して、ルート剛体部品の三次元位置を求める。カメラ方向での位置は既知なので、画像平面と平行な面上での位置が問題となる。

具体的には、部品位置決定木のルート剛体部品の中心軸の画像内の位置、及び画像と垂直な方向を軸とする回転角の計三つのパラメータを最初に求め、次いでルート剛体部品の中心軸の回りの回転パラメータを求める。制約条件から、手指とカメラとの距離は一定であり、その方向での位置を求める必要はない。それゆえ、スケーリング問題は生じない。

ルート剛体部品の位置決定法を以下で説明する。

1. まず、シルエットが画像の端と交わる点から、その輪郭に沿って直線検出を行う。
2. これをシルエットの両側について行い、得られた二本の側線間の中央線を求める。これがルート剛体部品の中心軸方向となる。
3. 両側線の画像中央よりの端点の中点を、ルート剛体部品の中心軸の端と

する (図 5) 。

4. 求めたシルエット上の中心軸に、ルート剛体部品の中心軸の投影が一致するように始めの三つのパラメータを計算する。
5. 手指モデルでは、ルート剛体部品は楕円錐台であるので、その投影像の中心軸に沿った輪郭は直線である。これがシルエットの両側線に一致するように回転パラメータを計算する。

step 5 では、実際には楕円柱で近似計算し、四つの可能な解を得る。この手指モデルの位置に対する四つの解各々について、次節の形状推定を行う。

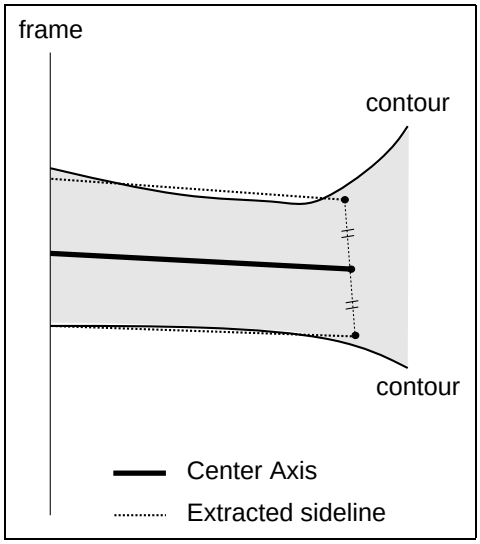

図 5: ルート剛体部品の位置決定

4.2 シルエットを利用した関節パラメータ決定

本節では、手指モデルを画像上に投影した領域とシルエットとの重畳情報をもとに、部品位置決定木を用いながら全ての剛体部品の位置を求めていく方法について述べる。剛体部品の位置は、関節パラメータを決定すれば定まるので、実際には関節パラメータを求めることになる。

前述のように、関節パラメータは 22 個あり、その全ての可動範囲を定義域としてシルエットに対する最適解を全解探索によって求めることは、非効率的である。本手法では、対象がシルエット画像であることから、面積を認識戦略の鍵とする。一般に画像認識では、画像から得られる局所的特徴が張る探索空間が巨大で、モデルの知識による探索空間の削減が非効率となることが多い。本研究では面積のみを三次元形状推定の鍵とするので、探索空間の大きさが抑制される。このような場合、構造的なモデリングを行っていれば、探索空間の絞り込みが効率的に行える。

実際の処理は、以下の手順で行われる。

＜関節パラメータの決定順序＞

関節パラメータの決定は、逐次的に行われ、どの関節から処理を行っていくかは部品位置決定木の根から葉への探索によって決定される。探索中に親ノードから子ノードへと処理が進む際に、そのアークに対応する関節パラメータが決定され、子ノードの剛体部品の位置が定まる。全てのノードが探索され、アークに付随する関節パラメータが決定されれば、処理が終了する。

＜各アークでの関節パラメータ決定＞

関節パラメータ決定の手順において、ノードの処理順が $n_0, n_1, \cdots$ となったときのノード i での処理を考える。ノード n_i の子ノードへのアークのうち、ノード n_j へのアーク j が選択されたとする。アーク j に対応する関節パラメータ $\mathbf{p}_j$ の可動範囲を一定の幅 w でサンプリングし、それらの中央値を $\mathbf{p}_j^k$ とする。投影関数 $P(j, \mathbf{p}_j^k, pix)$ は、ノード j の剛体部品を関節パラメータ $\mathbf{p}_j^k$ で投影した時に、画素 pix に投影が存在すれば 1 を、そうでないとき 0 を返す。また、シルエット関数 $S(i, pix)$ は、ノード i の処理時点において、画素 pix にシルエットが存在すれば 1 を、さもなくば 0 を返す。このと

き、

$$\min_k \sum_{pix}(S(i,pix) \oplus P(j,\mathbf{p}_j^k,pix))$$

を満たす関節パラメータ値 $\mathbf{p}_j^k$ を一時的な解とする。ここで、$\oplus$ は排他的論理和を表す演算子である。

さらにこの得られた中央値 $\mathbf{p}_j^k$ を含む定義域の幅 w を d 個に分割し、同様の処理を行って d 分探索を繰り返す。これを r 回実行して得られた $\mathbf{p}_j^{min}$ が、アーク j の関節パラメータの最終的な解となる。この二段階探索法は、必ずしも最適解を得られるとは限らないが、最適解に近い解を与えることが多いので、実質上問題はない。

同一のシルエット領域に複数の剛体部品が投影され、その一方でどの剛体部品の投影にも対応しないシルエット領域が残る可能性を排除するため、解 $\mathbf{p}_j^{min}$ が得られた直後に、次の処理をシルエット画像に対して行う。

$$S(n_{i+1},pix) = S(n_i,pix)\cdot(1-P(j,\mathbf{p}_j^{min},pix)),\ for\ all\ pix$$

これにより、各剛体部品は、それ以前にパラメータが決定した剛体部品の投影領域が覆い残したシルエット領域を覆うように、その関節パラメータが決定されることになる。

4.3 評価

本手法では、ルート剛体部品の位置決定時に解が四つ求まるので、その各々にシルエットを用いた関節パラメータ決定法が適用され、最終的な解も四つ求められる。この中で次の評価式の値が最も小さい解を、最終的な手指の三次元形状推定結果とする。

$$\frac{\sum_{pix}(S(n_0,pix) \oplus \bigcup_i P(i,\mathbf{p}_i^{min},pix))}{\sum_{pix} S(n_0,pix)}$$

$\bigcup_i P(i,\mathbf{p}_i^{min},pix)$ の項は、全ての剛体部品の投影の各画素位置での論理和を取ることを意味する。本式は、シルエット領域に対するモデルの投影とシルエット領域との不一致の割合を意味している。

5 実験及び考察

5.1 実験環境ならびに実験結果

図 4に対して、本手法を適用した結果が図 6及び図 7である。図 6は原画像と同じ位置パラメータで手指モデルを見たものであり、図 7はそれを前腕の中心軸のまわりで 115^o 回転させて見たものである。また、図 8は評価式を適用する際の、シルエットとモデルの投影との排他的論理和を表示している。この場合の評価値は 0.155 であり、これはシルエット領域に対する投影の不一致率が 15% ほどであることを意味する。

また、別のシルエット画像である図 9に対する実験結果が図 10であり、それを 60^o 回転させたのが図 11である。この実験での評価値は 0.261 であった。図 9と図 10の投影との排他的論理和を、図 12に示す。本推定結果では小指が折れ曲がっている。

なお、本実験で用いた制御変数は表 2の通りである。実験では、モデルのシルエット画像への投影に、透視投影 (perspective projection) ではなく、平行投影 (orthographic projection) を用いている。これは、実験で用いたカメラの焦点距離が $1500\mathrm{mm}$ なのに体して対象物体の奥行きが $100\mathrm{mm}$ 程度に過ぎないので、透視投影が平行投影で充分近似できるからである。部品位置決定木の探索には、深さ優先探索を用いた。

表 2: 関節パラメータ決定時の制御変数

初期サンプリング幅 (w)	20.0^o
二回目以降の分割数 (d)	4
反復回数 (r)	2

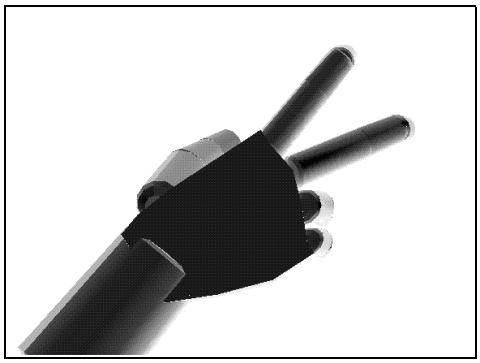

図 6: 推定された手指

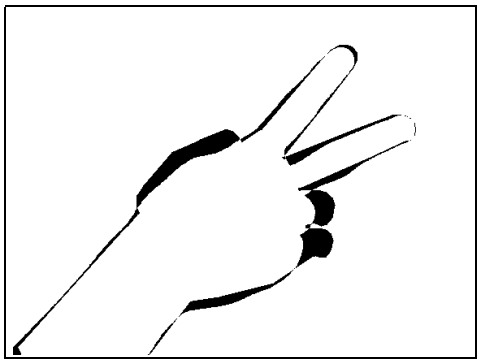

図 8: 入力データとの排他的論理和

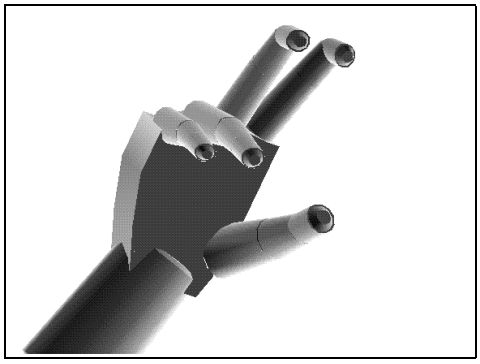

図 7: 推定された手指の 115^o 回転

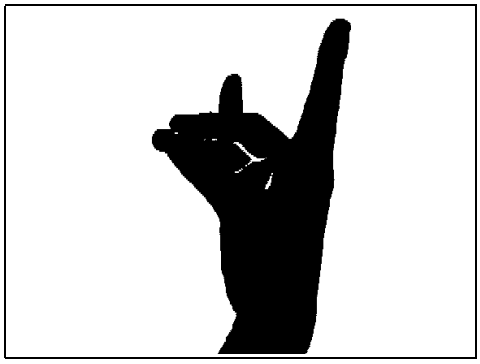

図 9: 入力シルエット：別例

5.2 考察

ここで利用した三次元モデルは、手指を対象として考案したものである。しかし、部品位置決定木で表現できるような変形可能な物体に対しては、本研究方法はそのまま適用できると考えられる。例えば、人体も、腕、足、胴、頭などに分割すれば、本手法がそのまま適用出来る。各剛体部品の表現方法に制約がないというのも、応用を行う上で有利である。

一方、現在の手法には、大きく分けて二つの問題点がある。

一つ目は、元の輝度画像と一致しなくても、良好な評価値をもつ結果が得られてしまうことである。部品位置決定木の探索順序が異なれば、解も変化する。これは、画像情報の制限を提案し、シルエット画像を用いている以上、三次元世界と二次元のシルエットの世界とは多対一の関係にあるので、このこと自体は不可避の結論である。しかし、同一シルエットに対する可能なパラメータ値の集合といった形で、シルエット画像からの一対多の解を求めることが望まれる。

二つ目は、部品位置決定木の構造上の問題である。木構造であるので、親ノードの結果が子ノードに影響を及ぼす。すなわち、一旦親ノードが局所的な解に陥ってしまうと、現在のアルゴリズムでは子ノードでの解が意味のないものとなる。これに対しては、子ノードの剛体部品の投影に対する評価値を導入し、その評価値が低ければ親ノードの解を改善する枠組が適切であろう。こ

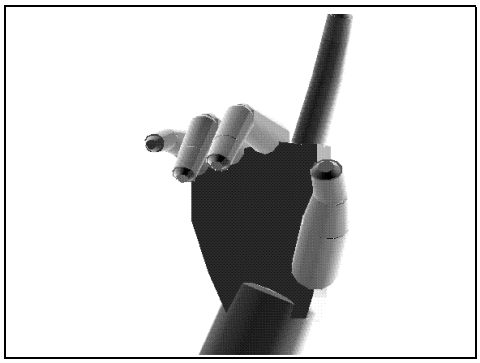

図 10: 推定された手指：別例

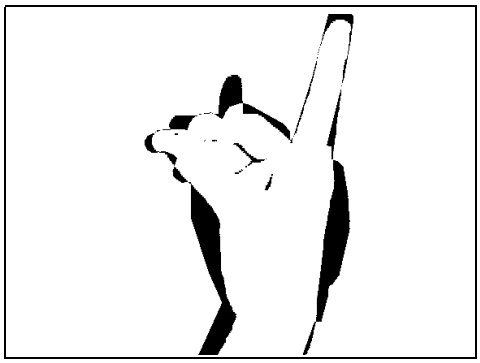

図 12: 入力データとの排他的論理和：別例

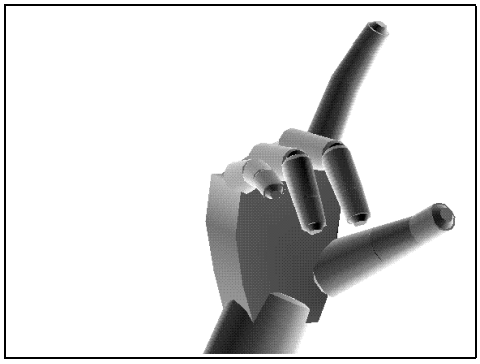

図 11: 推定された手指の 60^o 回転：別例

の場合でも、処理量の増加を抑える必要性がある。

6 おわりに

本手法は、

1. 手指の三次元形状推定時に面積を鍵とすることで、画像からのボトムアップ的情報の量を抑える。
2. 部品位置決定木という、比較的汎用性の高いモデル構造記述法を用いる。

という二つの方法を採用することで、手指のような変形可能な物体を対象とする三次元形状推定法を提案した。

しかし、1. の条件ゆえに、ほぼ同じ評価値の結果が出ても、人間の主観から見ると納得し難い状況が生じ得る。これは、人間が三次元形状推定を行う際に鍵としているのが、面積だけではないことに由来している。如何に探索空間の広がりを抑制しながら、様々な画像情報を三次元形状推定の戦略に加えていくかが、今後の研究課題となる。次に加える情報には、輪郭線情報を考えている。

参考文献

[1] R.A.Brooks, "Symbolic reasoning among 3-D models and 2-D images," AI, vol. 17, pp. 285-348, 1981.

[2] 久野, 岡本, 岡田, "認識戦略作成と位置決めに置ける対象モデルの利用," 信学技報 PRU 90-52, vol. 90, no. 124, pp. 85-92, July 1990.

[3] H. Rijpkema, M. Girard, "Computer Animation of Knowledge-Based Human Grasping," SIGGRAPH, Computer Graphics, vol. 25, no. 4, July 1991.

[4] J. Ponse, A. Hoogs, D. Kreigman, "On Using CAD Models to Compute the Pose of Curved 3D Objects," CVGIP Image Understanding, vol. 55, no. 2, pp. 184-197, March 1992.