

3次元ボクセルデータに基づく人体の姿勢推定

村上 裕介, 松山 隆司

京都大学大学院情報学研究科知能情報学専攻

概要 ボクセルの集合として表現された, 時系列3次元人体形状データの姿勢を推定するため, 多関節表現した人体構造モデルを, ボクセルデータにフィッティングするアルゴリズムを提案する. 人体構造の各パーツを楕円筒体積モデルで表現し, 身体の中心から端に向かってフィッティング結果を伝搬させることで, 計算量が少なく, 精度, 安定性に優れたフィッティングを実現した. また, 実験を通じて, 本手法が実時間で動作することを実証するとともに, アルゴリズムの有効性と問題点を検証する.

Estimation of the Human Pose from 3D Voxel Data

Yuusuke MURAKAMI, Takashi MATSUYAMA

Department of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto Univ.

abstract To estimate the human pose from a sequence of 3D voxels of a moving human, I propose an efficient algorithm that fits a multi-joint human body structure model to voxel data. I create the model composed of elliptical cylinders and fit the model to voxel data starting a center of the body and spreading the result towards the extremities. The realized fitting algorithm is speedy, accurate, and stable. I show that this algorithm can work in real time, and inspect the efficiency and problems of this algorithm.

1 序論

実世界のシーンを視覚的に記録するためのデバイスとして, カメラがよく用いられている. カメラにより, 手軽かつ高速に実世界の映像を記録することができる. しかし, この様にして得られる記録は, 3次元の情報に2次元の画像に縮退するために, シーン中の物体の3次元位置情報や, カメラの視点から遮蔽される部分の情報失われてしまう. 3次元の実世界を, そのまま3次元のデータとして取り込むことができれば, このような問題は解決できる. そこで, より忠実に実世界を記述するため, カメラで撮影された2次元画像から, 如何にしてシーンあるいは物体の3次元情報を獲得するかという研究が行われてきた. その中でも特に, 多視点映像に基づく3次元ビデオ [1] は, 実世界をより忠実に記述するメディアとして期待されている. これは, 対象を囲むように配置されたカメラで撮影した多視点映像から, 対象の全周囲3次元形状を復元し, 実写映像から得られる対象表面の色, テクスチャ情報を,

復元した3次元形状に張り合わせることによって, 実シーンの3次元情報をほとんど失うことなく記録できるメディアである. 3次元ビデオの特徴として, 閲覧者は視点位置と視線方向を指定することにより, 3次元データで構成される対象を, 自由な方向から見た場合の映像を得ることができる.

本研究の目的は, 人の3次元形状データから, コンピューターが自動的に人体各部の3次元位置を求め, 全身の姿勢を推定することである. これにより, 3次元ビデオで, 特定の部分を追跡した映像なども, 容易に記録することができるようになる. 更に, 姿勢の時系列変化がわかれば, 動作の認識などに研究に応用していくことが期待でき, 人とコンピュータのインタラクションを考える上で, 非常に有用な研究である. そのため, 本研究で提案する姿勢推定アルゴリズムは, このような応用研究も視野に入れ, 時系列で変化する実データに対して, リアルタイムで姿勢推定が出来るように, 高速性を第一に考え, 同時に姿勢推定の正確さと時系列的な安定性を追求するものとする.

姿勢推定の方法として、本研究では次のようなフィッティングによる方法 [2] をとる。

step1 対象人物の全身 3 次元形状データを得る。

step2 人体を単純な形状の組合せで表現したモデルを設定する。

step3 データとモデルをフィッティングさせることにより、モデルの姿勢から、データの姿勢を間接的に推定する。

本研究ではこの様な方法で姿勢推定を行う上で、モデルの設定とフィッティングのアルゴリズムについて、計算量と頑健性の両面から効率的な方法を提案する。

以下に本論文の構成を示す。第 2 章ではフィッティングの対象となる 3 次元ボクセルデータについて述べる。第 3 章ではデータに当てはめる人体モデルの構成について述べる。第 4 章ではフィッティングの具体的な手法について述べ、第 5 章で実空間上で姿勢の推定を行った実験、およびその結果を紹介し、第 6 章で結論を述べる。

2 3次元ボクセルデータ

本研究では、フィッティングの対象データとして、人体の全周囲 3 次元形状をボクセルの集合で表現したものをを用いた。ボクセルとは、3 次元座標がある物体によって占められているかを、1/0 の 2 値で表現したものである。ボクセルを用いると、対象形状は 3 次元座標のリストで表されるため、3 次元の並進、回転の座標変換にも容易に対応できる。

形状データの取得は視体積交差法 [3] により行なった。視体積交差法では、複数のカメラを用いて撮影された多視点映像から、カメラ毎に対象シルエットを抽出し、カメラ中心とシルエットにより定義される錐体の全論理積を対象形状として復元する。得られた対象の 3 次元形状の内部にボクセルを配置することで、対象形状がボクセルの集合として表されたデータが取得できる。しかし、この様にして得られるボクセルの集合は、必ずしも完全に対象形状を再現しているわけではなく、形状復元の手法上、実際の対象形状に対してボクセルの欠落や余分なボクセルが生じるなどの問題を含む。(図 1 参照) また、今回使用した形状復元システムでは、システムの解像

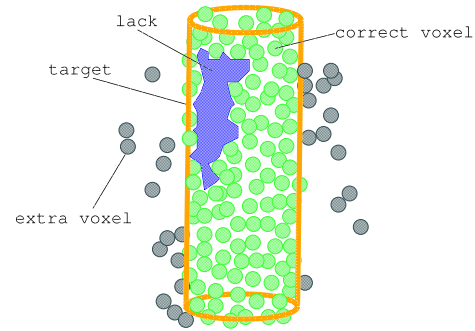


図 1: 対象と得られるボクセルの関係

度の関係上、あまり細かい部分の形状は正確には再現できないため、人体の細かい部分、手先や足先、鼻や耳などは形状がボクセルデータにほとんど反映されない。これらの問題は、ボクセルデータを使って姿勢推定を行う上で、推定の誤りなどの悪影響を及ぼす恐れがある。

3 人体モデルの設定

本章では、ボクセルで表現された人体形状データに対して、フィッティングが高速かつ正確に行なえるモデルの構成について考察する。

3.1 モデル全体の構成

本研究で実現しようとしている機能は、人の姿勢を解析するのに必要な、人の形状モデルを定義し、人体ボクセルデータに対してフィッティングを行なうことである。また、2 節で述べたように、ボクセルデータは、システムの解像度不足により、対象人物の詳細な形状を再現できないため、モデルは頭、胴体、腕、足などの、大まかな人体構造が表現できれば良い。この様に、人の構造を大まかに表現するモデルとして、人体各部を表す基本形状が、互いに関節で結合した多関節モデル [4] がある。本研究では、人体の姿勢を十分表現できるよう、全身を頭、胸、腰、左右の上腕、下腕、上腿、下腿の 11 個の基本形状 (パーツ) に分ける。

パーツ間は関節で接続されているので、隣接するパーツ位置には従属関係が考えられる [5]。この関係は、あるパーツを根ノードとして、その他全てのパーツをノードとして含む木構造で表し、親ノード

の位置が変化すると、子ノードの位置も同時に変化する、と考えることで容易に表すことができる。

また、全身のボクセルの重心は、ほぼ腰の位置に来ることが実験的に確かめられる [4]。これより、腰を従属関係の根ノードとし、姿勢推定は位置の推定が容易な腰部から始め、腰に近いパーツから順番にフィッティングし、パーツに付随する関節位置を逐次的に決定することで、推定の自由度を減らし、効率的に行うことができる。

3.2 各パーツの表現

この節では、効率的なパーツの表現法を考察する。個々のパーツの表現法には以下の表現が考えられる。

1. 軸指向モデル: 各体節パーツを線分で表現する。
2. 体積指向モデル: 各体節パーツを体積を持った塊で表現する
3. 表面指向モデル: 各体節パーツを表面を持った領域で表現する。

そこで、2章で述べたような不完全なボクセルデータに対してフィッティングする際の、評価法と推定結果の信頼性について、1つずつ検証するために、図1のようなボクセル群に、各モデルをフィッティングさせる場合を考える。図中の円筒は、実際に対象が存在する領域、緑の点は正確に形状を表すボクセル、灰色の点は余分なボクセル、青く描かれた領域はボクセルの欠落が生じているものとする。

3.2.1 軸指向モデル

軸指向モデルでは、モデルを表す線分と各ボクセルの距離の総和を評価値とし、これを最小化することでフィッティングを行なう。この場合、評価値を求める計算量は比較的少ない。一方、ボクセルの欠落や冗長なボクセルが軸に対して偏って発生すると、図2の様に推定の精度が低くなる。更に、このモデルの大きな欠点は、各パーツ単体として見た場合、軸方向の回転を表せないことである。

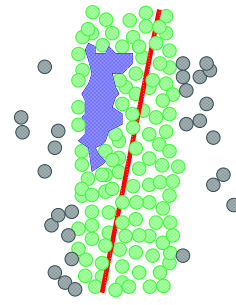


図 2: 軸指向モデルでの予想推定結果

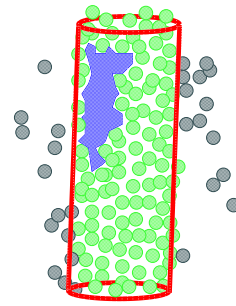


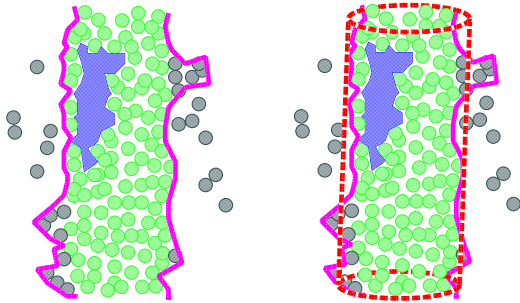
図 3: 体積指向モデルでの予想推定結果

3.2.2 体積指向モデル

体積指向モデルでは、モデル体積内のボクセル数を評価値として最大化することでフィッティングを行なう。欠落や余分なボクセルがモデル境界面から離れている場合、影響を受けないので、軸指向のモデルより信頼性が高い。推定結果は図3の様になる。一方、評価値の計算量は、各ボクセルのパーツ領域に対する内外判定の繰り返しになり、パーツの形状に依存する。

3.2.3 表面指向モデル

全身ボクセルデータから、図4の様に、対象の表面形状を再現すれば、その表面とモデル表面の最小距離の総和を評価値とし、これを最小化することでフィッティングを行なえる。再現される表面形状に滑らかさなどの拘束を与えることで、ボクセルの欠落や冗長なボクセルに対しては、その影響を抑えることができ、推定結果は図4の様になる。しかし、表面形状の再現というステップを余分に必要とし、評価値の計算では、有限な2つの平面間の最小距離を何度も計算することになり、計算量が大きいとい



ボクセルからの表面形成 予想推定結果

図 4: 表面指向モデル

う欠点を持つ.

以上の考察から, 評価値の計算量と推定結果の信頼性の両面を考慮し, 体積指向で各パーツを表現する.

3.3 多関節楕円筒モデル

これまでの考察から, 本研究では, 体積指向で個々のパーツを表現した 11 個のパーツにから成る多関節モデルを採用する. (図 5 参照) フィッティングの評価値はパーツ領域内に入ったボクセルの数とする.

さらに, 各ボクセルのパーツ領域に対する内外判定をより高速化するため, 全てのパーツは楕円柱で表現する. (この楕円柱の軸を以後パーツ軸と呼ぶ) これにより, パーツ軸が z 軸と重なるような座標系に, ボクセルの座標を座標変換することで, 非常に高速に内外判定が行なえる. さらに, パーツ間には関節の接続関係により, 中央 (腰) から末端へ向かう従属関係を定義する. (図 6 参照)

実際の人間には関節の可動範囲に制限があるが, このモデルには制限を設けない. さらに, 各パーツは体積を持っているが, パーツ同士は重なることが出来ないなどの制限も設けていない. すなわちフィッティング時パーツ軸は, 一方の関節を回転の中心として, あらゆる方向に向くことができる. このようにした理由は, 1 つはモデルの関節の表現を単純化する為であり, もう 1 つは関節の曲がる方向に制限を設けると, 探索不能となる恐れがあるからである.

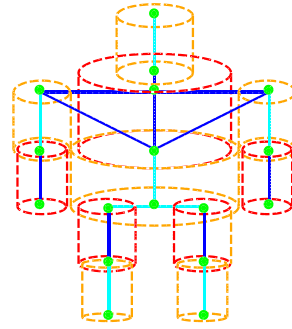


図 5: モデルの構成 (模式図)

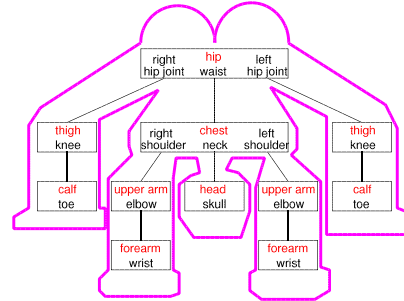


図 6: 各パーツ位置の従属関係

4 姿勢推定

この章では, 姿勢推定アルゴリズムを, フィッティングの評価法, 個別のパーツについての軸方向の決定法, 全体の流れ, の順で説明する.

4.1 評価基準

3 章の考察より, モデルの領域内に存在するボクセルの数を評価値に採用する. すなわち, 評価値 F は総ボクセル数を n として

$$F = \sum_{k=1}^n g(x_k y_k z_k) \quad (1)$$

と表される. ただし g はボクセル座標がパーツ領域内のとき 1, それ以外のとき 0 となる 2 値関数である. F を最大化するのがフィッティングの目的である.

4.2 パーツ軸方向の推定

本節では個々のパーツについて, パーツ軸方向を推定する方法を説明する.

まず、推定の自由度について説明する。今回はフィッティング時の各パーツのサイズは固定とする。パーツ内に含まれる関節のうち、既にフィッティング済のパーツと繋がっている関節を固定し、そこを中心として、パーツ軸方向を変化させ、最も評価値が大きくなる方向にパーツ軸を向ける。すなわち、パーツ軸の方向のみの2自由度の探索を行なう。

次に軸方向の更新方法であるが、単純な山登り法などの最適化手法を用いて、パーツを動かしては、別のパーツに属するボクセル群 (図8中ピンクの点) に対してもフィッティングしてしまい、容易に局所解に陥ることが考えられる。また、推定の高速度の観点からも、パーツ軸方向の更新は、出来る限り狭い範囲のボクセルだけを用いて、無駄な探索をせず、パーツ軸の更新方向を決定したい。関節以外の場所で人体が曲がることはないので、あるパーツの一方の関節位置がデータとモデルでほぼ一致していれば、その関節から確実に、まっすぐある方向にパーツの長さだけボクセルが分布している筈である。この方向を正確かつ高速に調べれば良い。そこで、次のようなフィッティング方法を提案する。

探索の初期状態は、図7で表す。青で描かれている線分がパーツ軸である。

step1 現在パーツに入っているボクセルの数 N_{now} を数える。

step2 パーツを軸方向に垂直に p 分割し、各領域内 $R_i (1 < i < p)$ のボクセル数 $N_i (1 < i < p)$ を数える。

step3 定数 k をカットオフとして定め、可動中心側から順番に $N_i < k$ となるかを調べる。

step4 最初に $N_i < k$ となる領域の、1つ前の領域 R_{i-1} 内のボクセルの重心 G を求める。

step5 パーツ軸を G を通る方向に更新した場合、パーツ全体に入っているボクセルの数 N_{next} を数える。

step6 $N_{next} > N_{now}$ ならばパーツ軸を更新し step1 に戻る。 $N_{next} < N_{now}$ ならば現在の軸を推定結果として探索を終える。

ここで、 k はパーツの底面の面積 S と定数 $a (0 < a < 1)$ を用いて $k = Sa$ となる定数である。

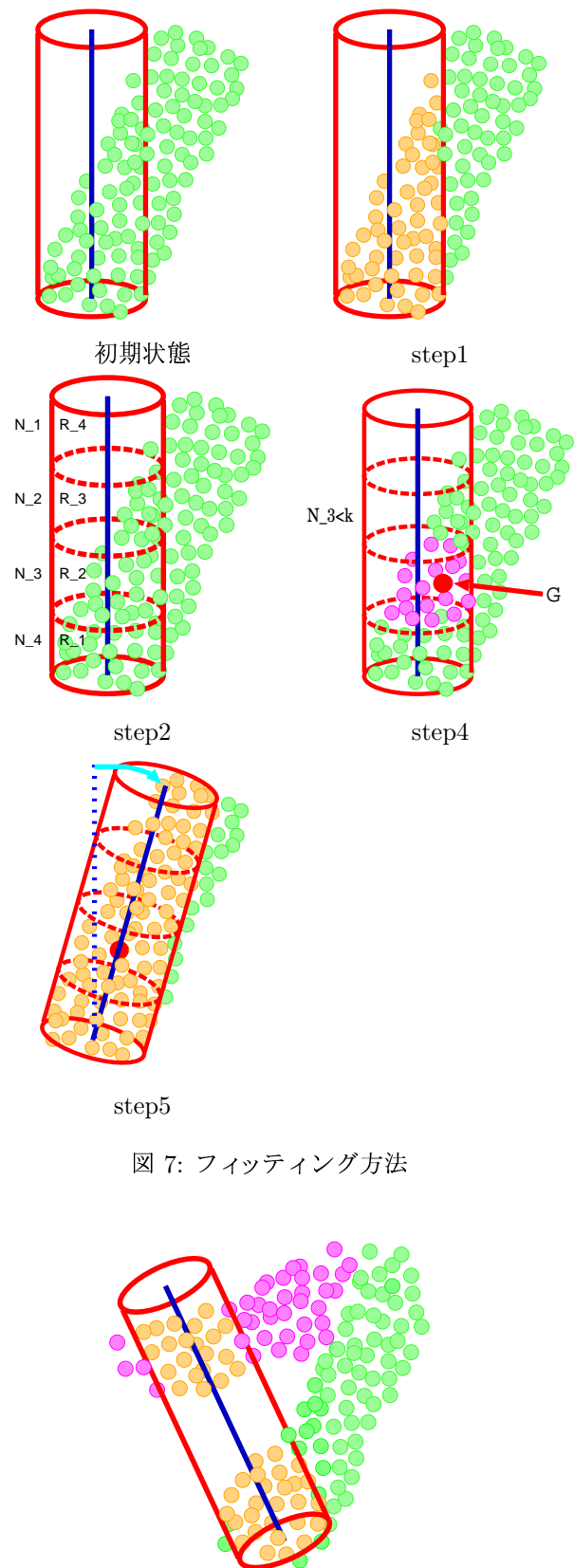


図 7: フィッティング方法

図 8: 局所解に陥った場合

更新するパーツ軸方向が決定すると、現在の軸を推定された方向に回転する。このとき、フィッティング中のパーツに従属するパーツの軸全てを回転する。これにより、パーツの相対的な位置関係は維持される。次に、他の推定に悪影響しないよう、フィッティングされたパーツに含まれるボクセルを取り除く。以上の作業で、1つのパーツのフィッティングは完了し、そのパーツに隣接する関節の座標は確定する。

このアルゴリズムは基本的には、山登り法的に評価値を上昇させるが、人体ボクセルデータの特徴を利用して、更新方向を探索無しで決定し、推定の高速化を計っている。また、回転中心から空間的に連続していないボクセルは軸の更新に影響しないため、図8のような局所解に陥る可能性が、単純な山登り法を用いた場合より少ないと考える。ただし、この方法が有効であるための条件として

条件 1 モデルの回転中心が、データの関節位置とほぼ合っている

条件 2 異なる部分のボクセル群同士が完全に一体化していない

の2つが満たされる必要がある。条件1はモデルの正確性と、推定そのものの正確性に依存する問題である。一方、条件2は、たとえ満たされなくとも、体積の大きなパーツから先にフィッティングするなどの工夫をすることで、ある程度克服できると考える。

4.3 姿勢推定の流れ

姿勢推定は図6の従属関係の根ノードである腰部から始めて、子ノード側へ逐次的に行なう。これは、動き易さとパーツの体積の大小の両面から考えて妥当である。そこで問題になるのは、最も基本となる腰の関節位置をどのようにして決定するかである。単純に考えると腰部の探索の自由度は、並進3回転3の6自由度となる。実際には全身のボクセルの分布から、もう少し手がかりが得られそうであるが、非常に困難な探索である。今回は立っている人間を対象に限定することで、腰部の探索を次のように行っている。

全身のボクセルの重心を計算すると、多少の手足の動きがあっても、ほぼ腰のあたりに来ることが実

験的に確かめられるので、重心座標に一定の補正を加えた座標を腰の関節位置とする。これで腰部のフィッティングが可能になる。

次に左右の上腿、下腿の順にフィッティングを行い、胸、頭、左右の上腕、下腕の順で逐次的にフィッティングを行い、全身の姿勢推定が完了する。

以上をまとめると、姿勢推定の手順は以下のようになる。

step1 全身ボクセルデータの重心から、腰の位置を決定する。

step2 腰部の軸方向を推定する。推定は4.2節で述べた手法で行う。

step3 既に推定の終わったパーツに繋がっているパーツの推定を行う。

step4 全てのパーツの軸の推定が終るまで、step3を繰り返す。

5 実験

これまでに述べた姿勢推定の枠組みを検証するために、アルゴリズムを実装し実験を行う。

5.1 実験詳細

計算機 CPU Pentium III 633MHz, Main Memory 320MB : 1台

対象データ 入力となるボクセルデータは身長170cm、体重60kgの男性の時系列形状データであり、フレームレートは毎秒約9枚である。ボクセルは全て幅2cmの格子点状に並んでおり、1フレームあたりの総ボクセル数は平均約10000点である。

使用モデル モデルは3章で説明した多関節楕円柱モデルである。各パーツのサイズおよび関節位置は、実際に対象となった人物の身体を直接測定することでおおよそ決定しているが、より良くフィッティングするよう実測値を基に調整している。モデルの初期値は図9に示す。1フレーム目はこれを初期値とし、2フレーム目以降は直前の推定結果を初期値としている。表示

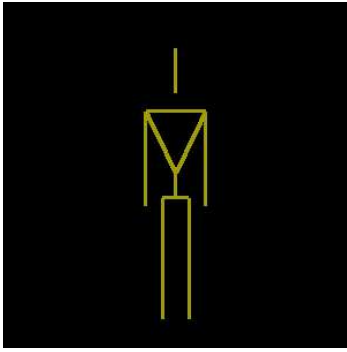


図 9: モデルの初期値

は各パーツ軸のみを表示している。(図 5 参照, 胸部については, 両肩と腰を繋ぐラインを表示している)

5.2 推定結果

入力したボクセルデータと姿勢推定した結果を並べて表示する。

姿勢推定にかかる平均時間は約 0.02 秒で, モデル領域内の総ボクセル数は平均約 9300 点であった。これは 90%以上のボクセルがモデルに含まれていることになる。推定結果(図 10)から, 姿勢がある程度正確に推定できていることがわかる。

5.3 考察

結果からわかる問題点を以下に述べる。

問題点 1 姿勢の変化に伴い, ボクセルの重心が多少移動するため, 腰の位置がずれ, 推定に大きく影響している。

問題点 2 パーツ軸方向の回転に対応出来ない。

問題点 3 あるフレームで大幅な推定誤りを起こすと, その後のフレームの姿勢推定に大きく影響し, 誤推定が連続する。

問題点 4 実際の人間とモデルの, 肩の構造が大きく異なり, 上腕の方向によっては, 肩の位置が合わず, 腕の推定が出来なくなる。

本手法は逐次的なアルゴリズムであるため, 問題 1 は重大な推定誤りを招く可能性が高い。今回は腰

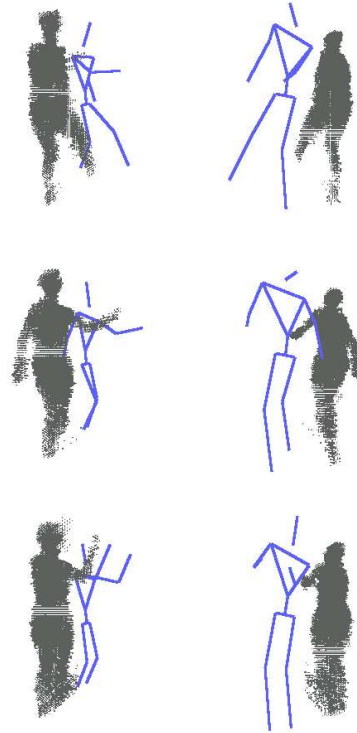


図 10: 推定結果

位置をボクセル重心位置のみを用いて決定しているため, 腰の推定位置がずれる原因となっている。重心や身長などの情報を利用して探索範囲をしぼる, あるいは, 上腿をフィッティング後, その情報を利用してもう一度腰の推定をやり直す, などの工夫をして, 腰位置の探索法を確立する必要がある。

問題 2 は, 今回のアルゴリズム上, 当然の結果である。パーツ軸探索後に軸方向の回転角の探索を行えば良いが, パーツ自体が扁平率の低い楕円柱であるため, 本研究で採用した評価法では, 評価値のピークが出にくく探索は困難である。軸方向の回転の探索は, 他の評価値を加えるなどの工夫が必要である。

問題 3 については, 探索アルゴリズムの信頼性を上げるとともに, 前のフレームの推定結果だけでなく, 前後の多くのフレームの情報を有効に利用することで, 明らかに誤った推定を取り除くなどの方法を考える必要がある。

問題 4 は, モデルの肩の構造を考え直す必要がある。

6 結論と今後の展開

6.1 結論

本研究では、ボクセルの集合で表された人体形状データから、高速にその姿勢を推定するために、人体各部を楕円筒体積指向で表現した多関節人体モデルを設定し、データに対して、身体の中心から末端に向かってフィッティング結果を伝搬させることで、効率的にフィッティングを行なうアルゴリズムを提案した。そして、実データに対して実際に姿勢推定を行い、提案したアルゴリズムが高速性、信頼性、安定性の面で、次のように有効であることを確認した。

高速性 1フレーム約0.02秒で推定でき、リアルタイムで動作した。

信頼性 図10より、姿勢解析としては十分正確に推定できている。ただし、5.4節の問題点2を解決しなければならない。

安定性 5.4節の問題点1,3,4により、動作の解析に使えるだけの安定性はない、特に問題4を解決しなければ、腕の推定は非常に不安定である。

6.2 今後の展開

今後の展開としては、実験で明らかになった問題を解決するとともに、さらなる展開として、以下の方法も検討したい。

6.2.1 入力データの加工

本研究で提案した姿勢推定手法は、入力されるボクセルを全て利用する。しかし、全てのボクセルを使わなくても、姿勢の推定は可能であると考えられる。例えば、明らかに孤立しているボクセルは冗長であり、これをあらかじめ取り除くことで、推定の精度だけでなく、データ量の減少による高速化も期待できる。また、多重解像度解析の枠組みを用いて、最初は荒い空間解像度のボクセルデータで、大まかな姿勢を推定し、徐々に解像度を上げて姿勢の絞込みを行っていくというやり方も有効であると思われる。

6.2.2 初期値の推定

姿勢推定を行う際、直前の推定結果をそのまま初期値とするのではなく、直前までに推定された姿勢の時系列の変化から、あらかじめ初期値そのものを予測しておく。これにより、より正解に近い初期姿勢から推定が可能になり、頑健性と高速性の両面で、パフォーマンスの向上が期待できる。

本研究は科学研究費補助金基盤研究 A 13308017 および特定領域研究 13224051 の補助を受けて行った。また、論文の執筆に当たっては、鷺見和彦先生に熱心な御指導を頂きありがとうございました。

参考文献

- [1] 松山隆司, 高井勇志, ウ小軍, 延原章平. “3次元ビデオの撮影・編集・表示”, TVRSJ Vol.7 No.4 pp.521-532, 2002.
- [2] A. Bottino, A. Laurentini. “A Silhouette Based Technique for the Reconstruction of Human Movement”, Computer Vision and Image Understanding No. 83, pp.79-95, 2001.
- [3] ウ小軍, 和田俊和, 東海彰吾, 松山隆司. “平面透視投影を用いた並列視体積交差法”, 情報処学会論文誌, Vol.42, No.SIG6(CVIM2), pp.33-43, 2000.
- [4] 美濃 導彦. “3次元人体モデル中心処理”, 情報処学会論文誌 Vol.43, No.SIG4(CVIM4), pp.10-24, 2002.
- [5] 西山 学. “3次元形状モデルを用いたカメラワークプランニング”, CVIM-133, pp.49-56, 2002.