

HMD に搭載可能なキャリブレーションレスモーショントラッカー

芹澤 隆史[†] 柳田 康幸[†]

A Calibration-less Motion Tracker Installable on HMD

TAKAFUMI SERIZAWA[†] YASUYUKI YANAGIDA[‡]

1. はじめに

現在、全身モーショントラッキングの用途がオフライン入力から、オンラインインタラクションに変化してきている。姿勢推定に関しては既に多数の研究や製品化が行われているが、既存手法では視界内の人物を全てキャラクターで上書きし HMD に表示する事は困難である。このような用途には、移動体に搭載可能なキャリブレーションレスのモーショントラッカーが必要であり、以下の要件を全て満たす必要がある。

- 全身の姿勢推定が可能
- マーカーレス
- カメラの自由移動可能
- ユーザーの体形、服装等の情報を必要としない

既存システムの大半は、事前登録されたユーザーが固定されたカメラの前で動作することを想定しているため、要件の大半を満たさない。また、複数のカメラを使用したスタジオ形式のキャリブレーションレスの姿勢推定システムや、距離画像を使用した移動体に搭載可能な姿勢推定システムの研究も行われているが、上記の全ての要件を満たすシステムは存在しない。

そこで、我々は距離画像カメラやステレオカメラから共通して得られるカラー画像と距離情報から、身体のエッジのみを優先的に抽出する技術と、身体のエッジをロバストに追跡する技術を利用し、三次元の姿勢推定を行う手法を提案する。また、提案手法を利用した姿勢推定システムを実際に開発し、当システムが上記の要件を全て満たした上で、キャリブレーションレスのモーショントラッカーとして 10fps で動作する事を確認した。

本手法の具体的な応用分野を以下に記す。

- 既存のモーションキャプチャーのユーザーキャリブレーションの簡便化。
- 自走型ロボットの非言語コミュニケーションの入力インタフェースとしての使用。
- 空港などでの不審動作の監視・分類など。

2. 従来研究

現在、多数のモーションキャプチャーに関する研究が行われている。本章では、全身・マーカーレスの姿勢推定手法とその特徴について述べる。

計測対象の周囲に配置した複数のカメラ画像から visual hull[1]等によってユーザーの身体モデルを復元し、人体の共通構造を使用することで姿勢推定を行う手法が Chi-Wei らによって研究されている[2]。この手法を用いれば、事前にユーザーの個別情報を入力する必要の無いシステムの実現が可能であるが、一方で、計測対象の全体をカメラで囲う必要があるため、移動体への搭載が困難である。

また、ステレオカメラや距離画像カメラにより計測対象に対して一方向からの距離イメージを取得し、事前に登録された計測対象の身体情報とカラー・距離画像の特徴点・ボリュームマッチングにより姿勢推定を行う手法が Schrotter G らによって研究されている[3]。これらの手法を用いれば、移動体に搭載可能な姿勢推定システムが可能となる。また、計測対象の身体情報の取得を自動的に取得するシステムも存在するが、計測には「カメラから一定距離、一定ポーズ、カメラの方向を向いて一定時間静止」のような制限が存在する。

自由空間・自由姿勢でのユーザーキャリブレーションが可能となり計測対象の身体情報を自動的に収集することが可能であれば、ステレオカメラや距離画像カメラを使用した姿勢推定手法が使用可能となる。

自由空間・自由姿勢でのユーザーキャリブレーションの研究としては、平尾公男らが腕領域の肌色を細線

[†] 名城大学理工学研究科

Meijo University Faculty of Science and Technology

[‡] 名城大学情報工学科

Meijo University Department of Information Technology

化する事で前腕・上腕の長さを推定する研究を行っている[4]。この手法は姿勢推定と同時にユーザーキャリブレーションを行うことが可能であるが、一方で、腕領域の肌色を認識しているため長袖の計測対象には使用できず、マーカースレスとは言い難い。

以上のように、姿勢推定に関して多数の研究が行われているが、キャリブレーションレスモーショントラッカーの要件を全て満たしたシステムは存在しない。

キャリブレーションレスのモーショントラッカーの実現が困難である理由の一つとして胴体の正面方向の推定が困難であることが挙げられる。

従来の全身・マーカースレスタイプの姿勢推定を事前情報無しで運用する場合、最初にユーザーの顔を検出し、体幹、上腕と太腿、前腕と下腿の順で方向ベクトルと回転を推定する。それぞれのボーンの基点は、親ボーン方向と回転に依存するため、親ボーン姿勢の推定精度が子ボーン姿勢推定に大きな影響を与える。

特に、胴体の正面方向の推定誤差は肩関節と股関節の位置に直接影響するため、手足の姿勢推定の過程で誤差の伝播・拡大が生じる可能性がある(図1)。しかし、システムが体形や服装の事前情報を持たず、一方向からのみの画像・距離情報から胴体の正面方向を推定する事は非常に困難である。

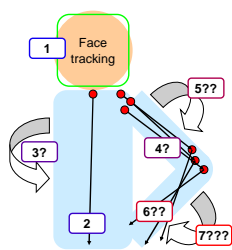


図1 誤差の蓄積

3. 提案手法

我々はソフトな関節を使用し、あえてボーン回転角を推定しない姿勢推定のアルゴリズムを提案する。

具体的な手法として、顔認識を行った後、胴体エッジの追跡により、胴体の方向と始点を求める。次に肩と股関節の位置を概算し、手足のエッジの追跡により、手足の方向と始点を求める。胴体の回転角が必要であれば、手足の始点から逆算する。

本手法の特徴は、ハードな関節構造を使用し、頭から姿勢推定を順次行う従来手法と比較して、推定誤差が子ボーンに伝播しない事である(図2)。

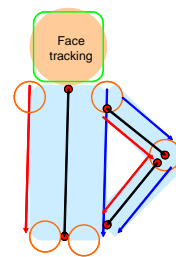


図2 提案手法

提案手法の実現には距離画像カメラやステレオカメラから共通して得られるカラー画像と距離情報から、身体のエッジのみを優先的に抽出する技術と、身体のエッジをロボラストに追跡する技術が必要である。

人体の各パーツ(胴体・上腕・前腕・太腿・下腿)は円柱に近似可能であり、円柱のシルエットには二本のエッジが必ず存在することを利用する。

以下に身体パーツの姿勢推定の概要を示す。最初にステレオカメラ、または距離画像カメラから、カラー画像と距離情報を取得する。次に、得られた画像からエッジとエッジの深度を抽出する。最後に顔の位置や親ボーンから概算された身体パーツの探索開始点付近からエッジの対を検出することで身体各部のボーン方向と角度の推定が可能となる(図3)。

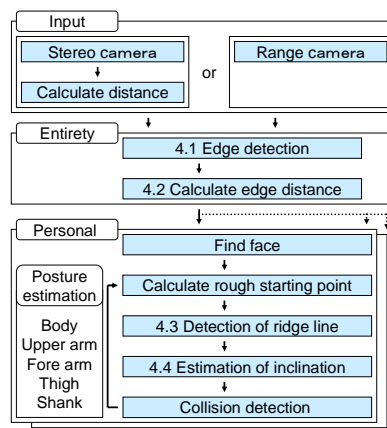


図3 概要

4. アルゴリズム

4.1 エッジの検出

提案手法では身体のエッジを追跡することで姿勢推定を行う。このため、背景の色と身体の色の変化の大小に関わらず安定してエッジを検出する手法が必要である。今回使用した手法を以下に述べる。

最初に、ガウスフィルタを使用し入力画像のノイズ成分を除去する。次に、エッジの判定を行うピクセルと周辺の8ピクセルの色を比較し、左右・左下-右上・上下・左上-右下の勾配を求める。最も強い勾配

と、強い勾配に直交する勾配の比率を計算し、比率が一定以上であり、周囲と比較して勾配が最も急であれば有効なエッジとして保存する（主勾配：距離画像の領域判別に使用）。有効なエッジであれば、最も強い勾配と隣接する勾配の比率からエッジの方向を12方向に細分化し保存する（副勾配：稜線の検出に使用）閾値を使用せずに明確なエッジのみを求めることで動画でのエッジ検出を安定して行うことが可能になった。結果を以下に示す（図4）。

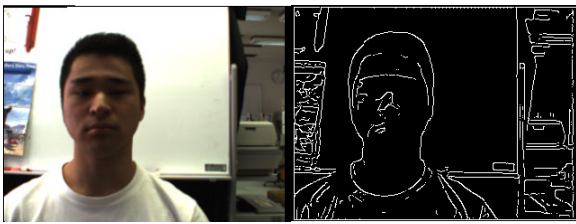


図4 エッジの検出 左：入力 右：主力

4.2 距離画像の領域判別

姿勢推定で使用する画像の領域は、身体側・エッジ・背景側に分割されるが、本手法で必要とされる距離情報はエッジの身体側のみである。

エッジの身体側・背景側の判別は、探索開始点が決定した時点で可能である。探索開始点から見て左側のエッジはエッジの右側から深度をサンプリングし、右側のエッジはエッジの左側から深度をサンプリングすることで、身体側の深度を取得することが可能である（図5）。サンプリング領域は4種類×9ピクセルのテーブルを用意し、主勾配の方向に応じて選択する。

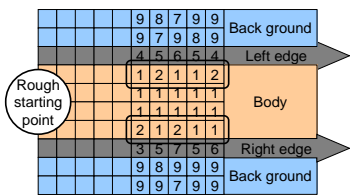


図5 領域判別

4.3 稜線の検出

提案手法では身体の稜線であるエッジの対を検出することで姿勢推定を行う。エッジの検出にはハフ変換が有効であるが、通常のハフ変換は一本のエッジを傾き0~ π の範囲で出力するため姿勢推定には適さない。我々はエッジの対を0~ 2π の方向で出力するようハフ変換を拡張した。今回使用した手法を以下に記す。

1. テーブルを利用して有効なエッジピクセルの位置を半径と接点に変換し半径と接点へ投票を行う（図6：右上，左上）。2. 半径と接点への投票結果から左右のエッジを決定する。3. 左右の得票を方向ごとに

集計し、最も有力な方向を求める（図6：左下）。4. 左右の輪郭線の基点の中間点を新たなボーンの基点として設定する（図6：右下）。である。

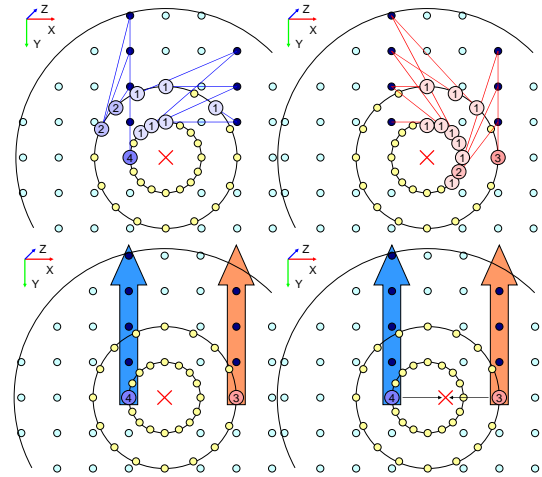


図6 稜線の検出

この処理により、始点のXY座標・Z回転が決まる。

4.4 前後の傾きの推定

三次元での姿勢を確定させるため、身体各部の前後の傾きや始点の位置を推定する必要がある。また、安定した姿勢推定を行うため、エッジ探索の失敗や深度推定の誤差等を考慮した上で、エッジが連続している場合、エッジが遮蔽される場合、エッジが途切れている場合を判別する必要がある（図7）。

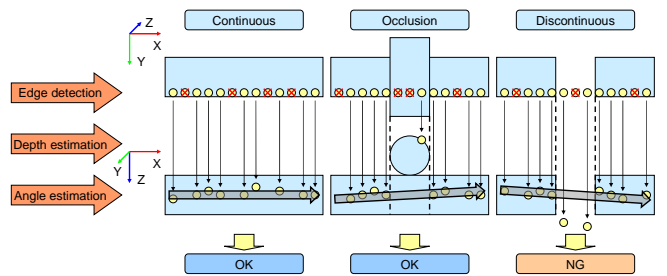


図7 エッジの分類

実装したシステムでは、遮蔽されたエッジを途切れたエッジより優先した上で傾きを決定するため、投票角度より手前の角度の投票対象に対してペナルティを課す方法を採用した（図8）

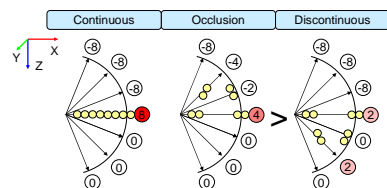


図8 奥行き情報の推定

この処理により、始点のZ座標・XY回転が決まる。

5. 結果

本システムは、ステレオカメラ入力と距離画像カメラの入力の両方に対応している。実際にベースラインが 6.5cm のステレオカメラ (図 9) (PointGrey Research Flea2; カラーVGA モデル) から画像を取得し、姿勢推定を行った。(CPU: インテル Core2Duo E8500 3.6GHz, RAM: 2GB)



図9 ステレオカメラ

システムの動作条件は、顔認識が成功すること、ユーザーがカメラから 3m~5m の範囲内に存在すること、計測対象の周囲 50cm 以内に障害物が存在しない事、以上の3項目のみである。

処理速度に関しては、320*240pix の画像を使用した場合、処理時間は平均 100ms+カメラからの入力遅延であった。また、処理時間の内訳は、エッジの抽出と分類に約 30ms, ステレオマッチングに約 60ms, 姿勢推定は一人当たり約 10ms であった。

図 (10, 11) は単純な背景で全身がカメラの視界内に入った場合の姿勢推定結果である。

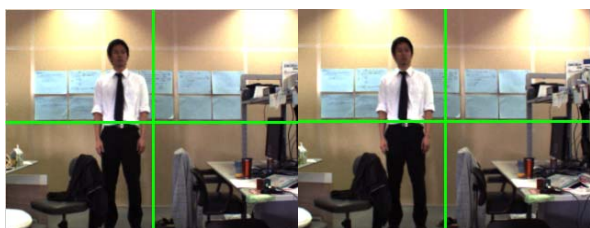


図10 入力1

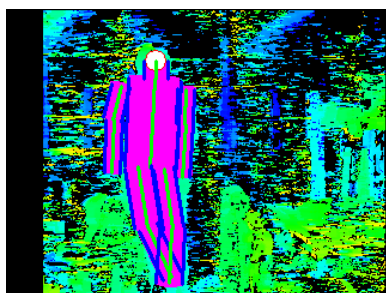


図11 結果1

図 (12, 13) は複雑な背景で身体の一部が画面外に出た場合の処理結果である。

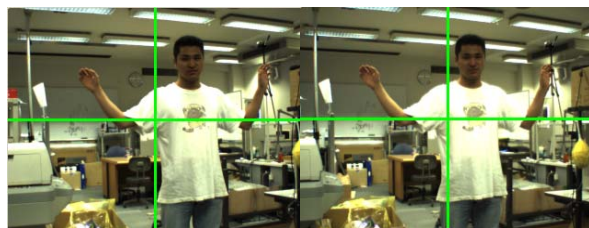


図12 入力2

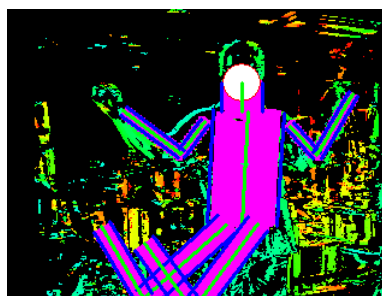


図13 結果2

6. まとめ

本研究では、カメラの自由移動が可能で、不特定のユーザーの姿勢推定が可能な姿勢推定手法を提案した。また、提案手法を使用した姿勢推定システムが準リアルタイムで動作することを確認した。

今後の課題はキャラクターの 3D モデルを実際に人体像にオーバーライドすることである。

参考文献

- 1) A. Laurentini. "The visual hull concept for silhouette-based image understanding". IEEE Trans. Pattern Analysis and Machine Intelligence. pp. 150-162, February 1994.
- 2) C. Chu, O C. Jenkins, and M J. Mataric. "Markerless Kinematic Model and Motion Capture from Volume Sequences". In Proceedings of IEEE Computer Vision and Pattern Recognition, Vol. 2, p. 475, June 2003.
- 3) G. Schrotter, "Markerless Tracking and Surface Measurements in Biomechanical Applications", International Conference on Robotics and Applications, Cambridge, MA, USA, October 31 - November 2, 2005
- 4) K. Hirao, A. Nakazawa, K. Kiyokawa, and H. Takemura. "Pose Estimation of Human Upper Body Using Multi-joint CG Model and Depth Images", Technical report of IEICE. PRMU, Vol. 104, No. 573, pp. 79-84, 2004