4M-9

インタラクティブ作品のための単視点モーションキャプチャ

齊藤貴志† 須藤智† 恩田憲一†
尚美学園大学芸術情報学部情報表現学科

1．はじめに

昨今、CG アニメーションの分野では、モーションキャプチャが普及している。モーションキャプチャは人のリアルな動きを取得できるので、ゲームなどの入力インタフェースに利用できれば、3D キャラクタとの高い一体感や作品の臨場感が得られると考えられる。しかし、システムを導入するにはきわめて高い費用が必要になる。これに対して、単眼視動画から姿勢を推定する研究がされているが、それらはリアルタイムでの処理を考慮していないため、ゲームのように高いレスポンスが要求される場面には適していない。

そこで、本論文ではリアルタイム処理のための人体姿勢推定手法として、人体のモデルの各部位全てを解析するのではなく、頭部、両手先足先の位置のみをテンプレートマッチングによって求め、それらをエフェクタ（effecor）とし、逆運動学（Inverse Kinematics）によって肘などの関節角度を求めることで、画像解析による処理速度への負担を軽減し、かつ 3D キャラクタの直感的な操作を実現する手法を提案する。

2．人体モデル

本手法では、図 1 のように人体モデルを定義する。モデルは腰をルートとするツリー構造で記述される。

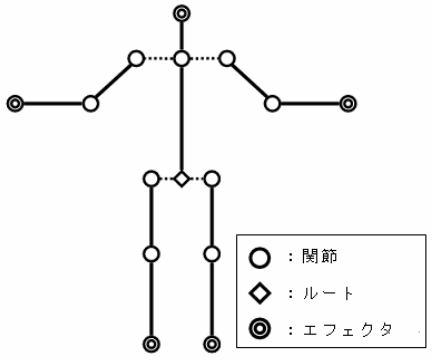


図 1　人体モデル

Mono-Viewpoint Motion Capture
for Interactive Applications

†SAITO Takashi　†SUDO Satoshi　†ONDA Norikazu

Shobi University
Faculty of Informatics for Arts
Department of Digital Expression

3．手法

本手法では、以下の処理手順で人体の姿勢推定を行う。

（1）前処理

姿勢推定処理を開始する前に、初期姿勢として図 2 のように T スタンスをとる被験者から、マッチング処理に用いるテンプレート画像を保持する。また、初期姿勢をとっている状態のシルエット画像から重心座標を求め、さらにそこから腰までの距離を加えたものをワールド座標取得における Y 座標最小限界値に設定する。



図 2　初期姿勢

（2）ワールド座標取得

体験者のシルエット画像から重心座標を求め、それをワールド座標とする。ただし、重心位置をモデルのワールド座標として利用する場合、現実の腰は動いていないはずにもかかわらず、検出結果の重心位置が動くことがあり得るため、得られた重心の X 座標を一定間隔で量子化し、不自然さを解消する。Y 座標は前処理で取得した最小限界値より値が低くならないように修正を加えることで、腰の位置が一定より上にならないようにする。

（3）テンプレートマッチングによるエフェクタ検出

エフェクタ位置の検出処理は、前処理で設定したテンプレート画像を用いたマッチングで行われ、マッチングによって得られた相関値が、あらかじめ経験的に定めておいたしきい値を最初に超えた領域の中心をエフェクタ位置とする。

（4）逆運動学による関節角度推定

逆運動学は、主にコンピュータアニメーションやロボット工学の分野で利用されるもので、関節物体の先端部分の位置を決めてから、各関節の角度を計算によって取得することができる方法である。

逆運動学の手法としては、代数的解法(algebraic solution)、ヤコビアン(Jacobian)を用いる反復による解法（iterative solution）などが挙げられるが、本手法では、一回の計算量が少なく安定した結果が得られ、かつ実装が容易なアルゴリズムであることから、発見的方法にあたる循環座標降下（CCD：cyclic coordinate descent）法を利用し、テンプレートマッチングによって検出されたエフェクタをもとに逆運動学を適応し、全身の姿勢を推定する。

ただし、逆運動学によって 3 次元の関節角度を計算する場合、解には多義性が生じるため、体験者の現実での姿勢と全く異なる姿勢が結果として得られる可能性がある。そこで、得られた関節角度から当該関節のワールド座標を計算し、それと、シルエット画像を照合し、適合しない場合は関節位置候補の中から適合する位置を探索することで、逆運動学によって生じる不自然さを解決する。

4．実験

まず、テンプレートマッチングによるエフェクタ検出実験を行った。その追跡成功例を図 3 に示す。なお、実験環境として、CPU クロック数 2.80GHz、メモリ容量１.00GB、画像処理ボードを搭載した PC を使用した。

サンプル動画像は 119 秒間(3540 フレーム)の入力に対して、処理を行った回数は 1330 回となった。よって、平均 FPS は約 11.18 となる。

また、各エフェクタの推定失敗数を計測したところ、表 1 のような結果が得られた。これらの結果から、テンプレートマッチングによる検出は有効であると言える。

表1　推定ミスしたフレーム数

エフェクタ	推定失敗（frame）
頭部	30
右手	279
左手	355
右足	82
左足	34

次に、エフェクタ検出からシルエット照合を利用した逆運動学の適応まで、手法の総合的な実験を行い、手法全体の有効性を検証した。

5．まとめ

本論文では、テンプレートマッチングを用いて頭部と両手先足先を追跡し、検出されたそれぞれの位置をエフェクタとし、肘や膝などの関節角度を逆運動学によって計算することで、単眼視動画像から人体の姿勢を推定するための方法を提案した。

本方法は、従来の研究のように全身の姿勢を全て動画像から解析するのではなく、逆運動学を適用するために必要となる 5 箇所のみを解析の対象とする。それにより、動画像解析で生じる処理速度への負担を軽減し、ゲームのように高速なレスポンスが要求される場面でも、インタラクティブ性が維持の有効性を検証した。

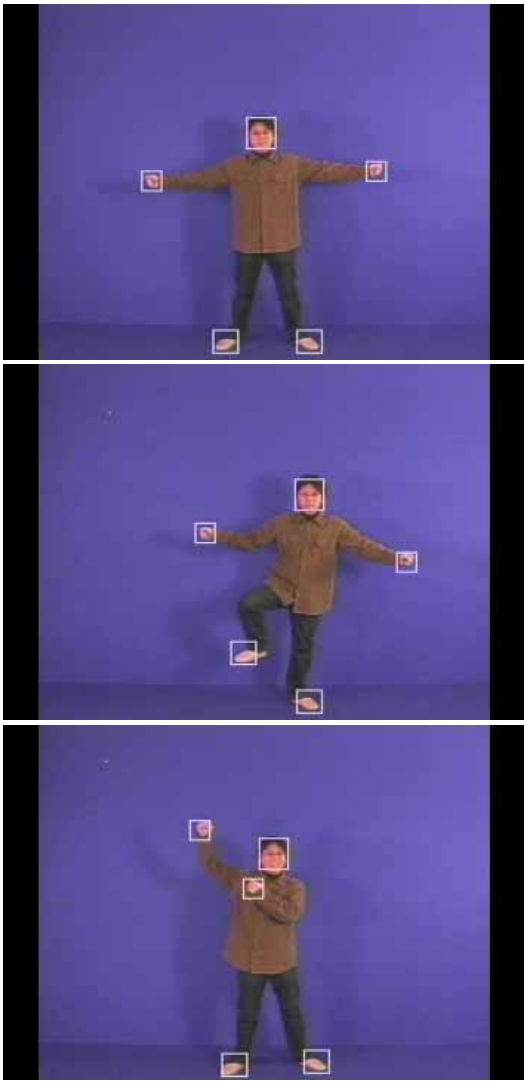


図3　成功例

今後の課題として、USB カメラでの検証が挙げられる。本論文の実験では画像処理プロセッサを用いることで、高速処理を実現したが、エンターテインメントへの応用を目的とするならば、より多くの人が楽しむためには個人で用意できる環境である必要がある。また、手法においては追跡精度を上げるためにヒストグラム特徴を利用することなどが考えられる。