

# 複数カメラによるリアルタイム・モデルベースモーショントラッキング

Real-time model-based motion tracking using multiple cameras

○ 森 敬充 (九州大)      正 倉爪 亮 (九州大)  
正 長谷川 勉 (九州大)      村上 剛司 (九州大)

Takamitsu MORI, Kyushu University, mori@irvs.is.kyushu-u.ac.jp  
Ryo KURAZUME, Kyushu University  
Tsutomu HASEGAWA, Kyushu University  
Kouji MURAKAMI, Kyushu University

This paper proposes a method for model-based motion tracking by multiple cameras using 2D-Distance-Field. The 2D-Distance-Field represents errors between the captured images and projected images of 3D model on the image plane. Using the 2D-Distance-Field, which is constructed rapidly using the Fast Marching Method, 3D motion of human is estimated robustly in real time. This paper shows some experimental results including simultaneous motion tracking of multiple people.

**Key Words:** Motion capture, Motion tracking, Multiple cameras, Level Set Method

## 1 はじめに

モデルベース・モーショントラッキングとは、人体の3次元モデルとカメラで撮影された画像上の人物像を比較することで、人間の動作を推定し、計算機内に取り込む手法である [1],[2]。本手法の多くは、カメラ画像と人体モデルの投影像との間で、エッジや輪郭線などの画像特徴量が最も良く一致する人体モデルの姿勢を探索し、関節角度などを決定する。しかし、最適な姿勢の探索は計算コストが大きく、リアルタイム処理の実現には計算量の削減が大きな課題であった。そこで本研究では、2次元距離場を用いて姿勢パラメータの修正量を直接求めることにより、リアルタイム処理が可能なモーショントラッキング法を提案する。

これまでに、C.Sminchisescu らによってシルエットベースの2次元距離場を用いた手法 [3] が提案されているが、本研究では2次元距離場を粗密かつ輪郭線ベースで構築することで、より高速かつ高精度なモーショントラッキングを実現する。さらに、複数カメラからの多視点画像を用いることで、自己遮蔽や光軸方向の運動にも頑強なシステムを構築し、複数人物動作の同時推定実験により有効性を検証する。

## 2 2次元距離場を用いたモーショントラッキングシステム

開発した手法は、背景差分により得られたカメラ画像上の人物の輪郭線に対して、2次元距離場を計算し、画像上に投影された3次元人体モデルの輪郭線上の距離値から姿勢修正量を求め、動作の推定を行う。システム構成を Fig.1 に示す。

### 2.1 3次元人体モデル

本研究で使用する3次元人体モデルは、Fig.2 に示すように、14個のリンクが13個の関節によって連結され、全21自由度を有している。なお、Fig.2 において座標系の数字は各関節での自由度を示している。ただし今回の実験では、3次元人体モデルの姿勢は、胴体中心の3次元位置と、首、手首、足首を除く8個の関節(各腕2、各脚2、全16自由度)の回転角度によって決定されるものとする。

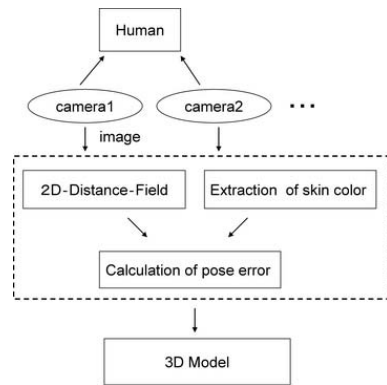


Fig.1: Proposed system

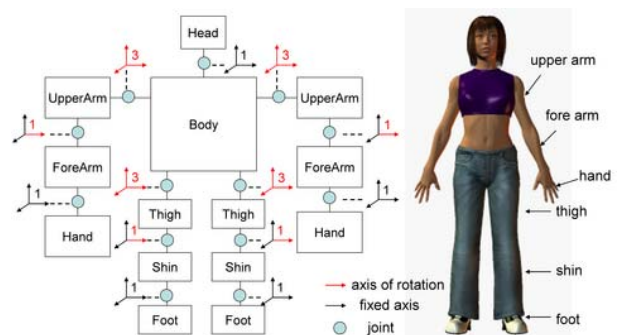


Fig.2: 3D body model

### 2.2 2次元距離場の構築

提案手法において、2次元距離場  $T$  とは、画面上の人物輪郭線  $I_h$  からの最短距離を画面上の全ての点で求めたものである。画面上の点  $p = (x_p, y_p)$  での  $T(p)$  は以下を満たす。

$$T(p) = 0, p \in I_h \quad (1)$$

$$|\nabla T(p)| = 1 \quad (2)$$

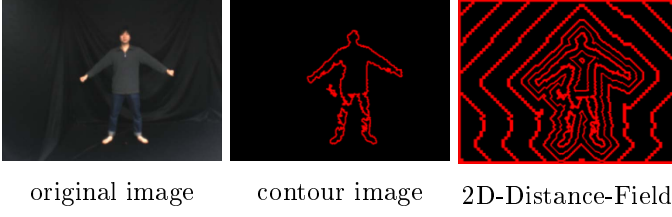


Fig.3: Construction of 2D-Distance-Field

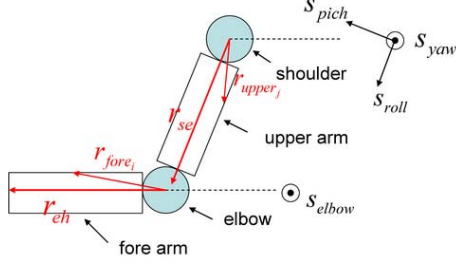


Fig.4: Arm model

具体的な2次元距離場の構築方法について述べる。まず、カメラ画像から背景差分によって人物シルエット画像を取得し、これにLevel Set Method[4]を適用して、人物輪郭線を抽出する。次に、抽出された輪郭線からの距離場  $T$  を Fast Marching Method[4]によって構築する。2次元距離場の構築の様子を Fig.3 に示す。ただし、本手法では、2次元距離場は輪郭線に近い領域のみ詳細に構築し、輪郭線から離れた領域では粗に計算することで、構築処理を高速化している。

次に、構築された2次元距離場を用いた3次元人体モデルの位置修正量の計算手順を示す。ただし、使用した3次元人体モデルは、いわゆるサーフェスモデルであり、表面は微小な多数の三角パッチにより構成されている。

まず、3次元人体モデルを2次元距離場に投影する。ここで、3次元人体モデルの投影輪郭線上の点  $p_i = (x_{p_i}, y_{p_i})$ 、及びそれに対応した三角パッチを  $u_i$  とする。このとき、点  $p_i$  に加えられる画面上の力  $f_i$ 、及び3次元人体モデル全体に加えられる画面上の力  $F$  を以下のように求める。

$$f_{p_i} = \begin{bmatrix} T(p_i)T_x(p_i)/D(p_i) \\ T(p_i)T_y(p_i)/D(p_i) \end{bmatrix} \quad (3)$$

$$D(p_i) = \sqrt{T_x(p_i)^2 + T_y(p_i)^2} \quad (4)$$

$$F = \sum_{p_i} f_{p_i} \quad (5)$$

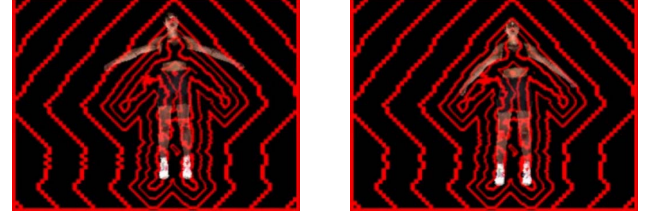
ただし、 $T_x$ 、 $T_y$  はそれぞれ  $T$  の  $x$   $y$  方向の一次微分値である。求められた  $F$  を3次元空間に逆投影し、3次元人体モデルの重心に加えることで、位置を修正する。

次に、各関節の修正量を求める。ここでは簡単のため、Fig.4に示す腕部分について考える。まず、3次元モデル前腕（手も含む）の投影輪郭線に対応するパッチを  $u_{fore_i}$ 、肘関節から  $u_{fore_i}$  へのベクトルを  $r_{fore_i}$ 、 $u_{fore_i}$  にかかる力を  $f'_{fore_i}$  とし、肘関節周りに発生するモーメントを

$$M_{fore} = \sum_{u_{fore_i}} (r_{fore_i} \times f'_{fore_i}) \quad (6)$$

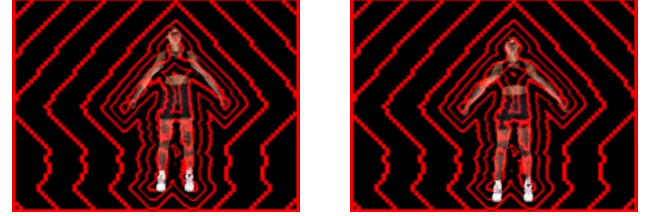
として求める。ここで、 $f'_{fore_i}$  は Eq.(3) で求められる画面上での力  $f_{fore_i}$  を3次元空間に逆投影したものである。これより、肘関節の回転軸方向を  $s_{elbow}$  とすると、肘関節の修正量は

$$\Delta\phi_{fore} = s_{elbow} \cdot M_{fore} \quad (7)$$



frame 0

frame 1



frame 2

frame 3

Fig.5: Examples of motion estimation

として求まる。

次に、3次元モデル上腕の投影輪郭線に対応するパッチを  $u_{upper_j}$ 、肩関節から  $u_{upper_j}$  へのベクトルを  $r_{upper_j}$ 、 $u_{upper_j}$  にかかる力を  $f'_{upper_j}$ 、肩関節から肘関節へのベクトルを  $r_{se}$  とする。これより、肩関節周りに発生するモーメントを

$$M_{upper} = \sum_{u_{upper_j}} (r_{upper_j} \times f'_{upper_j}) + \sum_{u_{fore_i}} ((r_{fore_i} + r_{se}) \times f'_{fore_i}) \quad (8)$$

として求める。さらに、肩関節の回転軸方向を  $s_{roll}, s_{pitch}, s_{yaw}$  とすると、肩関節の修正量は

$$\Delta\phi_{roll} = s_{roll} \cdot M_{upper} \quad (9)$$

$$\Delta\phi_{pitch} = s_{pitch} \cdot M_{upper} \quad (10)$$

$$\Delta\phi_{yaw} = s_{yaw} \cdot M_{upper} \quad (11)$$

として計算できる。

### 2.3 肌色情報の利用

手足が胴体などに重なって撮影された場合、本手法では手足の輪郭線を抽出することができず、2次元距離場だけでは姿勢修正量を正しく求めることができない。そこで提案手法では、2次元距離場に加えて手足の肌色情報も利用する。ここでも簡単のため、腕部分について考える。

まず、3次元モデルをカメラ画像上に投影し、3次元モデルの手先位置から画面上の肌色位置へのベクトルを  $f_{hand}$  として求め、これを3次元空間に逆投影した  $f'_{hand}$  を求める。次に、3次元モデルの肘関節から手先位置へのベクトルを  $r_{ch}$  とし、肘関節の修正量を以下のように求める。

$$M_{fore2} = r_{ch} \times f'_{hand} \quad (12)$$

$$\Delta\phi_{fore2} = s_{elbow} \cdot M_{fore2} \quad (13)$$

同様に、肩関節の修正量を以下のように求める。

$$M_{upper2} = (r_{se} + r_{ch}) \times f'_{hand} \quad (14)$$

$$\Delta\phi_{roll2} = s_{roll} \cdot M_{upper2} \quad (15)$$

$$\Delta\phi_{pitch2} = s_{pitch} \cdot M_{upper2} \quad (16)$$

$$\Delta\phi_{yaw2} = s_{yaw} \cdot M_{upper2} \quad (17)$$

以上、Eq.(7)、(9)~(11)、(13)、(15)~(17)を両腕両脚について求めることで、3次元人体モデルの姿勢を修正する。Fig.5に実際に姿勢を推定している様子を示す。

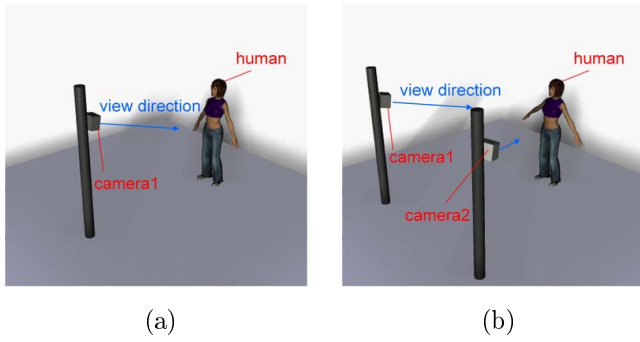


Fig.6: Experimental setup



Fig.7: Tracking results by a single camera

### 3 トラッキング実験

#### 3.1 単一カメラによるモーショントラッキング

まず、単一カメラによる歩行動作の推定実験を行った。なお、本実験はオンラインで推定を行っており、処理時間は8fps(Pentium D 3.00GHz)、使用したカメラはDragonfly2(Point Grey Inc.)である。実験環境を Fig.6(a) に、実験結果を Fig.7 に示す。このように、単一カメラでもある程度の動作推定は可能であるが、脚の奥行き方向(カメラ光軸方向)の推定が不正確なことが分かる。

#### 3.2 多視点画像によるモーショントラッキング

次に、複数カメラを用いて同様の歩行動作の推定実験を行った。実験環境を Fig.6(b) に、推定結果を Fig.8 に示す。複数カメラを用いることにより、3次元人体モデルの姿勢をより正確に推定できていることが分かる。

さらに、椅子に座る動作に対する実験結果を Fig.9 に示す。このように、脚関節を大きく曲げる動作も推定可能であった。

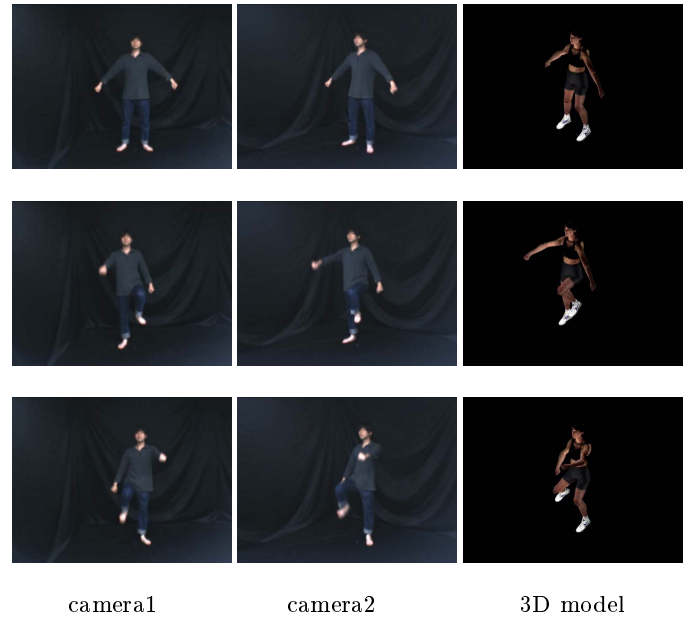


Fig.8: Tracking results by two cameras

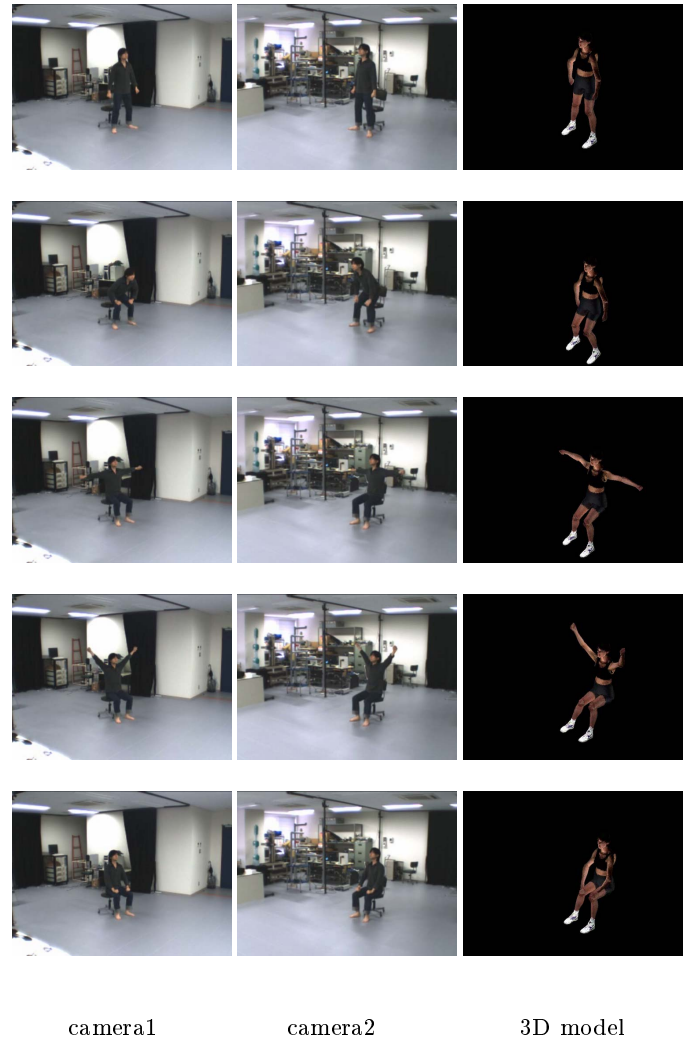


Fig.9: Tracking results for sitting motion



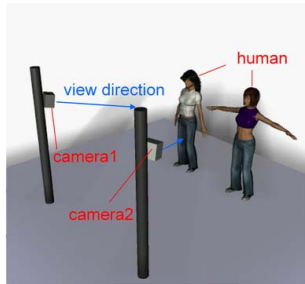


Fig.10: Experimental setup

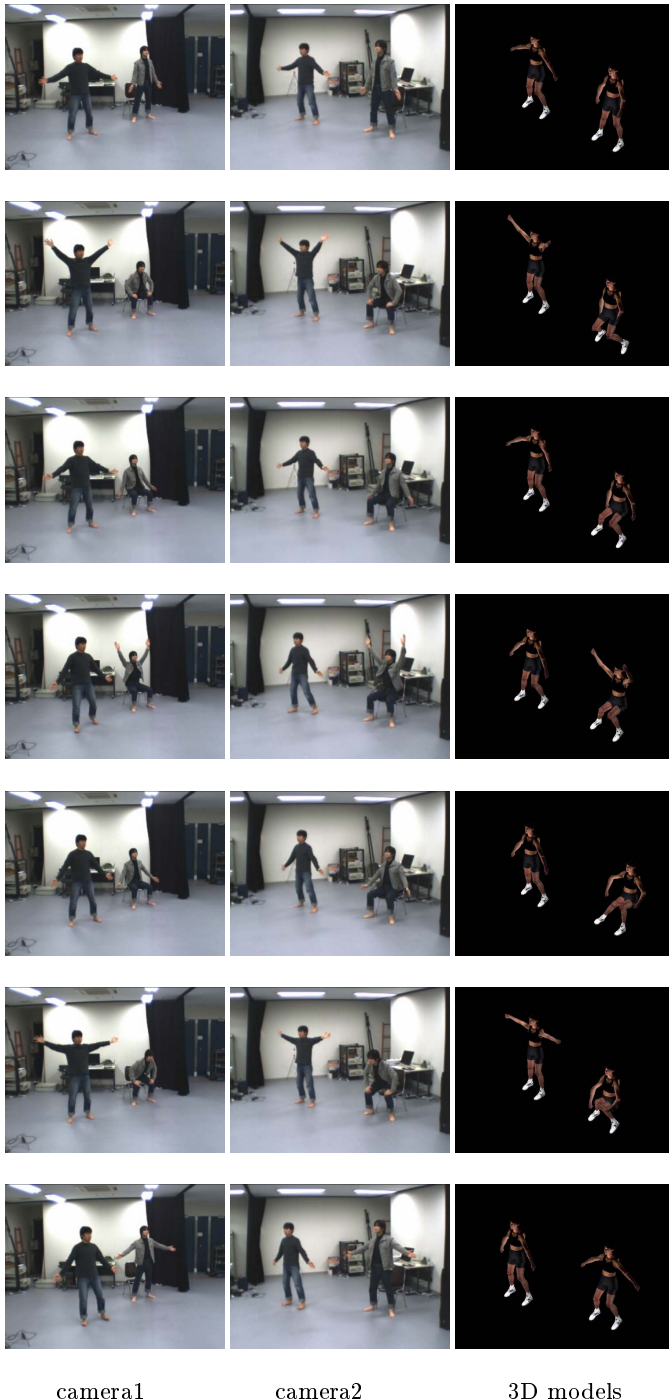


Fig.11: Tracking results of two people by two cameras

### 3.3 複数人物動作の同時推定

2人の人物動作を同時に推定する実験を行った．実験環境を Fig.10 に示す．また，実験結果を Fig.11 に示す．これより，2人の人物に対しても，1人の場合と同様に動作推定が可能であることを確認した．なお，提案手法は，一度2次元距離場を構築すれば，追跡する輪郭数が増えた場合でも，姿勢修正量の計算時間はほとんど変化しないという特徴があり，複数人物動作の同時推定に特に適していると考えられる．

## 4 まとめ

本稿では，輪郭線ベースの粗密2次元距離場を用いたリアルタイム・モーショントラッキング法を提案した．また実験により，歩行動作や椅子に座る動作などが推定可能であること，また複数人の動作推定にも適用可能であることを確認した．

今後の課題として，現在使用している3次元人体モデルでは，椅子に座るときの腰の屈伸など細かい動作を再現できない場合があり，人体モデルの胴体の自由度を増やすことが挙げられる．

## 文献

- [1] Quentin Delamarre and Olivier Faugeras: " 3D Articulated Models and Multiview Tracking with Physical Forces ", Computer Vision and Image Understanding, Volume 81, Number 3, March 2001, pp. 328-357(30)
- [2] Masanobu Yamamoto, Akitsugu Sato, Satoshi Kawada, Takuya Kondo and Yoshihiko Osaki: " Incremental Tracking of Human Actions from Multiple Views ", 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'98), 1998.
- [3] C. Sminchisescu and A. Telea: " Human Pose Estimation From Silhouettes: A Consistent Approach Using Distance Level Sets ", WSCG International Conference on Computer Graphics, Visualization and Computer Vision, 2002
- [4] J. Sethian: " Level Set Methods and Fast Marching Methods, second edition ", Cambridge University Press, UK, 1999
- [5] Ryo Kurazume, Hiroaki Omasa, Seiichi Uchida, Rinichiro Taniguchi, Tsutomu Hasegawa: " Embodied Proactive Human Interface "PICO-2" ", International Conference on Pattern Recognition, B04-0206, Aug 2006.
- [6] 倉爪 亮, 内田誠一, 谷口倫一郎, 長谷川勉: " プロアクティブインターフェースロボット "PICO-2" ", 第9回ロボティクスシンポジウム講演会予稿集, RS036, (2006).